Perceptual and acoustic gender differences in the speech of 4 1/2 - 5 1/2 year old children.

Thesis

UNRESTRICTED

PERCEPTUAL AND ACOUSTIC GENDER DIFFERENCES IN THE SPEECH OF 4½ - 5½ YEAR OLD CHILDREN

MORAY NAIRN M.A. (HONS)

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy OF THE OPEN UNIVERSITY

QUEEN MARGARET COLLEGE

APRIL 1997

Date of award: 1st May 1997

ACKNOWLEDGEMENTS

I wish to express my gratitude to a number of people who have contributed to the final product in a number of ways.

I am grateful to both of my supervisors, Dr. Janet Beck and Dr. Nigel Hewlett, for their consistent and excellent advice on every aspect of the thesis and for their willingness to talk over any and every issue which might arise. There are several sections which have been improved following discussions with them however any errors or omissions are, of course, my own.

The path through the thorny field of experimental statistics was eased for me immeasurably with the help of Dr. Ellen Bard at the University of Edinburgh. Her patient explanation of, not only the procedural implementations of complex tests using specialised software, but also the important theoretical concepts which underpin some of the issues encountered has made (relatively) light work of what otherwise would have taken many, many fevered attempts.

Thanks are also due to my family and friends for their support and in particular to Janet and Alice for keeping me going through thick and thin. In those distant, early days, Ron Motherwell provided not only sound technical experience but a source of many (mostly welcome) distractions for which I am grateful.

My biggest vote of thanks, however, is due to the head teachers, staff, parents and especially children of Corstorphine and Carrick Knowe Primary schools in Edinburgh. Without the co-operation of so many willing parents and the help and support of the teaching staff, I could not have carried out the most important task of data collection. The fact that an initial estimate of the pool of subjects grew from 40 to over 100 when I came into contact with the children testifies to the extremely warm welcome which I received and for which I am so very grateful.

ABSTRACT

The linguistic factors which identify a speaker as being either male or female are reasonably well understood and documented when we are considering adult speakers. Many of these factors become apparent at puberty when the sexes diverge along predictable anatomical and physiological paths. It might be expected, therefore, that prepubertal children should appear relatively undifferentiated in terms of gender and that young boys' and girls' speech should be sexually homogenous.

This study has confirmed, however, that adult listeners can correctly identify the sex of a prepubertal child from samples of speech. Results of the present study yielded correct identification rates which varied between 66% (using isolated vowels as the sample) and 76% (using sentences as the sample) - all of these rates were significantly greater than chance. Girls were shown to be better identified by listeners than boys and female listeners tended to be more accurate at identifying gender than male listeners.

During the acoustic phase of the study, a number of parameters were selected for measurement which were regarded as likely to be involved in the gender-identification process. Overall, there was a surprisingly large number of negative results, with only a very few parameters yielding significantly sex-different outcomes. No differences were found in Fo between the sexes and only 5 out of 18 comparisons of formant frequency showed significant differences. An investigation into vocal breathiness indicated that, on the basis of this parameter, children could be assigned into perceptual groupings ('most / least male-like') better than into biological sex groupings ('boy / girl'). The conclusion reached is that listeners may use different acoustic cues to identify children's sex from speech than adult's sex, alternatively or additionally, they may be able to focus their perceptive skills more finely on the small acoustic inequalities that exist.

The concept of gender-specific speech is discussed in a general commentary of the various influences exerted in the formation of gendered-identities.

Table of Contents

CHAPTER THREE - RESULTS

CHAPTER FOUR - DISCUSSION AND FURTHER ANALYSIS

List of Tables

List of Figures

Chapter One
Review of the Literature

1.0 INTRODUCTION

> "Three things you can be judged by - your voice, your face and your disposition."
>
> Ignas Bernstein (1978)

There appears to be information encoded in the speech signal, beyond the actual linguistic message, which is readily available to the listener. If the speaker is known to us, it is likely that we will be able to identify him or her from voice alone. Even if the speaker is a stranger, a listener is able to make judgements about their personality, gender, age, emotional state, socio-economic status and the condition of their health. Differentiation of gender between adult male and female voices is arguably the most salient of these extralinguistic vocal characteristics, and it is generally accepted that the fundamental frequency (Fo) of voicing (and its perceptual correlate, pitch) is the main cue that distinguishes the speaker's sex, (Coleman, 1971; 1976). The lower Fo of adult males is a consequence of secondary sexual dimorphism occurring at puberty, specifically, the relatively greater growth of the vocal folds and larynx in males compared to females.

It might at first appear that the only method of distinguishing boys and girls before puberty is a visual one - in general, boys look, dress and behave in a different style to girls. Broadly speaking, there is little apparent differences in the vocal quality of boys and girls prior to the 'voice break' of puberty which results in a drop in the pitch of the male voice of roughly one octave.

Research has revealed, however, the possibility that prepubertal boys and girls can be distinguished on the basis of recordings of their speech alone and that, after taking account of any gender differences in speech content (vocabulary etc.), boys and girls may also have characteristically different voices in terms of differences in glottal source, vocal tract filter and other parameters.

The vast majority of the relevant studies in the field of acoustic gender differences have been carried out in the U.S.A. and therefore using American children. These studies give rise to questions as to the status of the gender discrimination process, one of which is to what extent is it language- or culture-specific? One of the hypotheses tested in this work, then, holds that the ability of listeners to recognise the gender of prepubertal children is not limited to American subjects, but is a universal ability. Using Scottish listeners and children as experimental subjects for the first time, The study will investigate whether the perceptual ability of these adults is highly comparable to similar results from across the Atlantic and, indeed, around the world.

There are further issues relating to the patterns of gender identification results. Are boys better identified than girls? Do female judges identify gender more accurately than male judges? Is there an interaction with respect to identification rates between speaker and listener sex? Are individual listeners biased to respond 'male' or 'female'? Each of these questions are addressed and experimental evidence is brought to bear in the search for answers.

Assuming a measurable gender difference in the perception of the boys and girls, a further aim of the study is to isolate one or more acoustic parameters which form the basis for this discriminatory ability. Previous research has been equivocal on the determinants of the vocal dichotomy of the sexes prior to puberty. The influence of a few important acoustic parameters is assessed in detail - these specifically are those parameters which are the likeliest candidates to underlie a gender distinction given what we know of adult vocal sex-differences. In addition, a number of 'lower profile' acoustic parameters, some of which have not been previously directly implicated in acoustic gender studies, are measured and compared. A full statistical analysis of the perceptual and acoustic results is included.

The remainder of this chapter addresses relevant topics in the field and considers the existing experimental evidence. Chapter 2 describes the experimental design of the present investigation and chapter 3 reports the results. In chapter 4 there is a discussion of the findings and conclusions are drawn in the light of the evidence.

The truth of the quotation at the start of this section will now be examined: can we really judge an individual by voice alone?

1.1 GENDER DIFFERENCES IN NORMAL ADULT VOICES

1.1.1. Introduction

The subject of the acoustics of the human voice, drawing as it does on such diverse fields as anatomy, physiology, aerodynamics, physics and engineering theory, is an immense topic. Indeed, the number of textbooks and dissertations that exist on the subject is testimony to the abundance of relevant information that there is to be discussed. When we concentrate specifically on the acoustics of the child's voice, we must add another dimension to our discussion - that of paediatrics. The developing child differs in structure and function from the adult equally in terms of vocal anatomy as in terms of other physical and behavioural traits.

It is necessary, therefore, to limit the scale of any such discussion on grounds of practicality. The investigations presented in this thesis propose to highlight some perceptual and acoustic sex-differences in samples of children's phonations. The author is aware that periods of voiceless speech may also contain acoustic cues to gender, for example, in measures of voice onset time or in the spectra of voiceless fricatives, however there are a number of factors which suggest the use of phonatory material. Firstly, the contribution of phonation / voicing information to the recognition of gender in adults is clear - fundamental frequency and formant information provide the acoustic basis for a clear-cut perceptual distinction between the sexes (Childers and Wu, 1991). Any frequency-domain measurements, including the important spectral analysis of the glottal source, will require a period of voiced speech as input. Secondly, the relative ease with which acoustic information may be extracted from vowels makes them attractive as experimental stimuli. Furthermore, the immature speech development of 4-5 year old children often expresses itself by means of large variations in the accuracy and reliability of certain consonant articulations but vowel productions tend to be generally more stable. Finally, from the point of view of situating the results within a shared theoretical framework of gender-differences in pre-adolescents' speech, the use of mainly phonatory material is important. Whilst there remain a number of routes of investigation which involve voiceless samples of speech, as the role of a child's fundamental frequency is not yet firmly established in the process of perceptual gender recognition from voice, there is clearly more work to be done using voice source parameters.

In order to more fully understand the process of phonation, it is necessary to be familiar with the mechanisms that give rise to it, that is, the vocal organs and vocal tract. The following section therefore presents a brief and selective anatomy of those areas of laryngeal structure which are deemed to be crucial to a reasonable understanding of phonatory function. Where appropriate, sex-differences in the anatomy of the vocal tract will be pointed out and commented upon. The reader will notice that little discussion is included either of the musculature or the innervation of the vocal organs. There is no indication of any significant differences that exist between the sexes in normal children in these areas, furthermore, these areas are well documented in standard medical textbooks (e.g. Dickson and Maue-Dickson, 1982; Romanes, 1971).

1.1.2 Anatomy of the larynx

A great deal of the anatomical facts presented in this section is based on information drawn from a relatively small number of relevant medical textbooks, therefore, rather than disturbing the reader with continued, repeated references to the same works, I intend to pay credit here to those of my sources who have contributed the most in the way of background information.

For breadth of anatomical detail Pick and Howden (1977) remains unsurpassed; descriptions of the laryngeal structures and some diagrams were adapted from Romanes (1971), Dickson and Maue-Dickson (1982) and Beck (1988). Hirano

(1981, 1983) provided information on vocal fold structure. Finally, Tanner (1978) and Sinclair (1978) were the major sources of diagrams and data for growth rates.

Further reference will be made in the text to authors who have made specific contributions of particular relevance to a topic.

1.1.2.1 General Structure, Function and Position of the Larynx

The larynx forms the upper end of the windpipe, extending from the trachea to the pharynx and has evolved to serve three main functions: airway protection; pressure-valving and phonation.

The larynx is surrounded on three sides, posteriorly and laterally, by the pharynx. The passages followed by air to the lungs and food to the stomach cross at the level of the larynx - the larynx empties into the trachea whilst the pharynx empties into the oesophagus. The larynx is therefore situated in the anterior section of the pharynx with the trachea lying anterior to the oesophagus.

The hyoid bone is roughly U-shaped with the points of the U to the rear and is a free-floating bone situated in the neck between the root of the tongue and the **thyroid cartilage**. The larynx, in turn, is suspended from the hyoid bone by means of membranes, ligaments and muscles. The walls of the larynx are supported by a number of cartilages including the **thyroid, cricoid and arytenoid cartilages.**

1.1.2.2 The Cartilages, Bones, Ligaments and Joints

<u>Thyroid Cartilage</u>

The thyroid is the largest of the laryngeal cartilages and is formed by the fusion of two cartilaginous plates (laminae). Only the inferior two-thirds of the anterior length of the cartilages fuse leaving the deep thyroid notch which separates the laminae superiorly. Immediately below the notch is the **laryngeal prominence** (also known as the 'Adam's apple') which is the most anterior part of the cartilage. In adult males the laryngeal prominence can often be clearly seen pressing against the tissue of the neck. The angle of fusion of the two thyroid laminae varies between 90°-120° and is smallest in the male which explains why the Adam's apple is more conspicuous in men than in women. It also accounts for the greater antero-posterior depth of the cartilage in the male.

The posterior margins of the thyroid cartilage taper superiorly and inferiorly into a pair of thin horns (cornua). The **superior horns** attach to the horns of the hyoid bone via the thyro-hyoid ligaments whilst the **inferior horns** articulate with the cricoid cartilage by the crico-thyroid articular facets.

When investigating measurements of the thyroid cartilage one finds that the length of the superior horns and the posterior separation of the laminae are relatively invariable, but many other proportions are prone to considerable variability, in particular sex-differences (Maue and Dickson, 1971). In a study carried out by Maue (1970) adult males showed the larger thyroid cartilage measurements for prominence of the Adam's apple, weight of the cartilage and overall size in the majority of cases. In the remaining cases there was no significant sex-differences; i.e. females never displayed greater values for thyroid dimensions than males. Furthermore, the line of fusion of the two thyroid laminae was significantly more rounded in females than in males.

Maue and Dickson (1971) cite some measurements of adult thyroid cartilages which appear to conform to general laryngeal measurements found in Gray's Anatomy (Pick, and Howden, 1977).

Table 1.1 Comparison of measurements of the thyroid cartilage and larynx from three sources

Maue and Dickson (1971)			Pick and Howden (1977)			Kaplan (1971)		
Thyroid cartilage measure	male (mm)	female (mm)	Larynx measure	male (mm)	female (mm)	Larynx measure	male (mm)	female (mm)
Height (from superior to inferior horn)	44	38	Vertical diameter	44	36	Vertical diameter (from upper epiglottis to lower cricoid)	70	48
Antero-posterior length	37	29	Antero-posterior diameter	36	26	Antero-posterior diameter	40	35
Weight	8g	4g	○	•	•	○	•	•
○	•	•	Transverse diameter	43	41	Transverse diameter	40	35
○	•	•	Circumf.	136	112	Circumf.	•	•

Cricoid Cartilage

The cricoid is smaller but thicker and stronger than the thyroid cartilage and is composed of hyaline cartilage. It forms the inferior and posterior parts of the cavity of the larynx and is situated immediately above the upper ring of the trachea to which it attaches by means of fibrous membrane. Occasionally the cricoid is seen to be fused with the tracheal cartilage either uni- or bilaterally. Although basically ring-shaped, the cartilage is enlarged posteriorly into a lamina which is roughly square when viewed from behind. The anterior part of the cartilage (the arch) is a narrow band which measures approximately one quarter to one fifth of the depth of the posterior lamina. The antero-posterior width, transverse length and laminar height of the cartilage are all of very similar dimensions in males and females.

The cricoid cartilage connects with two other cartilages: the thyroid and arytenoid cartilages. The cricothyroid articular facets are situated at the junction of the cricoid arch and lamina on either side and are the places of attachment of the inferior horns of the thyroid cartilage. The superior surface of the lamina of the cricoid cartilage is relatively short and terminates on both sides in a cricoarytenoid articular facet. These facets are angled downward laterally and anteriorly and articulate with the paired **arytenoid cartilages**. The cricothyroid and cricoarytenoid joints will be discussed in more detail below.

There are fewer obvious sex-differences in measurements of the cricoid cartilage than in measurements of the thyroid. Maue (1970) found once again that, where there was a significant sex-difference, the males were universally larger than the females in cricoid measurements. The average height of the cricoid lamina for adult males was 25mm and for adult females was 19mm. As was the case for the thyroid cartilage, the weight of the cricoid cartilage was twice as great in the male (5.8 g) than in the female (2.89 g).

Arytenoid Cartilages

The arytenoid cartilages are a pair of very small cartilages situated at the upper border of the cricoid cartilage at the back of the larynx. They are pyramidal in form consisting of three sides plus a base. The posterior (or dorsal) surface, which attaches to the arytenoid muscle, is triangular, concave and smooth and ends laterally in the posterior ridge. The medial (or internal) surface is narrow and smooth and is covered by mucous membrane. The plane of the medial surface forms the lateral boundary of the respiratory passage of the glottis. It projects anteriorly into the anterior ridge. The anterolateral surface is somewhat rougher and is delimited by the anterior and posterior ridges. The base of each arytenoid cartilage is concave and broad and is smoothed to allow contact with the cricoid cartilage. There are two basal projections worthy of note: firstly, the external angle, which is short, rounded and prominent, expands posteriorly and laterally and is called the **muscular process** (because the cricothyroid muscles attach to it); secondly, the base of the arytenoid extends anteriorly into the **vocal process**. The anterior ridge forms the superior border of the vocal process to which the vocal ligament is attached.

Maue (1970) found little individual variability in the dimensions of the arytenoid cartilages but there were some sex-differences which again indicated the larger structure in the adult male. Average arytenoid height in the male was 18 mm and 13 mm in the female. the average anteroposterior measure was 14 mm in the male and 10 mm in the female. Once again the weight ratio of the cartilage for males to females was 2:1 (0.39 g in the male and 0.2 g in the female).

Corniculate and Cuneiform Cartilages

The corniculate cartilages are two small cone-shaped nodules consisting of elastic fibro-cartilage which lie on the superior summits of the arytenoid cartilages and extends them posteriomedially.

The cuneiform cartilages are small conical cartilaginous bodies situated in the mucous membrane (aryepiglottic fold) which extends from the apex of the arytenoid cartilages to the **epiglottis**. There are normally at least two, and sometimes more, cuneiform cartilages.

Epiglottis (Epiglottic Cartilage)

The epiglottis is shaped like an inverted and flattened pear with the flattened surfaces facing anteriorly and posteriorly. Its structure is that of a rigid lamina of elastic fibro-cartilage and, lying behind the tongue and in front of the superior opening of the larynx, it forms the upper part of the anterior wall and the superior margin of the laryngeal cavity. The inferior point of the cartilage is attached to a point in the **thyroid cartilage** just below the thyroid notch by a long, narrow ligamentous band - the thyroepiglottic ligament. It is also connected to the posterior surface of the **hyoid bone** - the hyoepiglottic ligament.

The epiglottis is concave when viewed from the rear, i.e. it curves toward the tongue, and is covered with a mucous membrane. When this membrane is stripped away, the posterior surface can be seen to be studded with mucous glands situated in small pits.

Cricothyroid Joints

As was indicated above, the inferior horns of the thyroid cartilage link with the cricothyroid articular facets located on the lateral aspect of the cricoid cartilage to form a pair of synovial joints. The cricothyroid facets, unlike the remainder of the cartilage, are composed of soft tissue only and are thus specialized for their role as flexible pivots. The cricothyroid joints permit the cricoid cartilage to rotate about a transverse (i.e. horizontal) axis and hence the superior margin of the lamina of the cricoid cartilage and the attached arytenoid cartilages tilt towards or away from the anterior part of the thyroid cartilage. The vocal ligaments, which connect the arytenoid and thyroid cartilages, are tightened or slackened by this rotational movement and this changes the length of the **vocal folds**. Dickson and Dickson (1971) cited in Dickson and Maue-Dickson (1982) claimed that,

> "in manipulating fresh human larynges taken at autopsy, it has been found that by rotating the cricothyroid joint to its maximum degree, vocal fold length changes of approximately 25 per cent can be demonstrated This degree of change would seem to be consistent with that observed in the living larynx via cineradiography, photography and indirect laryngoscopy during pitch change".
> [p. 151]

<u>Cricoarytenoid Joints</u>

These joints are important in the understanding of phonation as they form the fulcrum of motion of the posterior ends of the vocal folds, which attach to the arytenoid cartilages. The joints themselves are small, synovial, saddle joints and are traditionally believed to allow two basic movements: the principal movement is an anterolateral-posteromedial sliding motion in which the arytenoid cartilages glide transversely on the articular facets of the cricoid cartilage so that they move closer together or further apart; another type of motion, rotation of each arytenoid about its vertical axis, swings the vocal process laterally and medially, thus separating or approximating the vocal ligaments. A number of studies of the cricorytenoid joint have contested the modes of movement involved, claiming that the rotational movement described above is not possible and that the sliding movement is very restricted (Fink, Basek and Epanchin, 1956; Cooley, 1964; Deweese and Saunders, 1973). Dickson and Maue-Dickson (1982) describe two modes of movement which occur at the cricoarytenoid joints; firstly, the familiar sliding motion, which these authors report to be severely limited by the presence of the tight, posterior capsule of the joint; secondly, a rocking motion about the short convex axis of the cricoid facet. The analogy that is drawn here is that of each arytenoid cartilage as a vertical beam attached anteriorly and posteriorly to the ground by two guy-wires which are the vocal ligament and the posterior cricoarytenoid ligaments respectively. Thus the beam (cartilage) is free to rock from left to right (laterally and medially) but not forward and back (anteriorly and posteriorly).

The Hyoid Bone

The hyoid bone is so named due to its resemblance to the Greek letter upsilon and is sometimes also called the lingual bone because it supports the tongue and is the point of attachment of many of its muscles. It lies between the root of the tongue and the thyroid cartilage and forms a moveable base for the tongue. The bone is shaped like a horseshoe with the curved arch to the front. It consists of a body and a pair of greater and lesser horns (cornua) which project backward from the lateral surfaces of the body. The greater horns are attached superiorly to the skull by the stylohyoid ligament and inferiorly to the thyroid cartilage by the thyrohyoid membrane.

1.1.2.3 The Vocal Folds

A) External Structure

The vocal folds are situated in the centre of the larynx and run from the inferior edge of the thyroid angle to the anterior part of the base of the arytenoid cartilages (the vocal processes). In the vertical dimension, the superior surface of the folds forms the inferior wall of the ventricle and the inferior surface extends downwards bilaterally to meet the superior surface of the cricoid cartilage. Each fold is roughly triangular in cross-section and consists of the upper conus elasticus, the vocal ligament and muscle fibres, the whole of which is covered by a mucous membrane.

For purposes of description it is possible to divide each vocal fold into two convenient sections based on the type of tissue present at the glottal edge. The anterior two-thirds of the length of each vocal fold where the vocal ligament borders the glottis can be referred to as the **ligamental** part of the fold, whereas the posterior one-third of the vocal fold length where the medial edge of the arytenoid cartilage forms the glottal border can be referred to as the **cartilaginous** part. This terminology may also be used to refer to parts of the glottis, for example, "the intercartilaginous glottis".

B) Internal Structure

Hirano (1975, 1977, 1981, 1983) and his colleagues (Hirano, Koike, Hirose and Morio, 1973; Kakita, Hirano, Kawasaki and Matsushita, 1976; Okada, 1978; Kurita, 1980; Hirano, Kurita and Nakashima, 1983; Kakita, Hirano and Ohmaru, 1981) have been responsible for investigating and demonstrating the multi-layered structure of the human vocal fold.

This is an important concept from several points of view. First, from the functional viewpoint, the different layers of tissue lend the vocal fold different mechanical properties as a vibrating body. Second, histologically, Hirano was able to demonstrate the stratification of the vocal fold structure into layers by describing the different cellular composition of each layer. Finally, from the clinical point of view, almost all pathologies of the vocal folds are believed to originate from a specific layer.

Hirano (1975) described the composition of the vocal fold as the mucous membrane (or mucosa) plus the muscle. He further divided the mucosa into **epithelium** and **lamina propria**. The final sub-division was of the lamina propria into three layers according to the density and composition of the fibres in each: the *superficial, intermediate and deep layers.*

The five different layers of vocal fold tissue are composed as follows:

- The epithelium. This is the outermost layer of the vocal fold and, as its name would suggest, is composed of non-keratinizing stratified squamous epithilial cells. Its function is to maintain the shape of the vocal fold and it is well suited to this task as it forms a relatively stiff, non-elastic layer of tissue which resists stretching.

- Lamina Propria (Superficial Layer). This layer consists of a network of loosely arranged elastic and collagen fibres which are embedded in a semi-fluid matrix. Hirano (1981: 5) regards this layer as "somewhat like a mass of soft gelatin". It is therefore likely to be the most flexible of the vocal fold layers.

- Lamina Propria (Intermediate Layer). Beck (1988) notes that there are many more fibres in this layer of the Lamina propria than in the previous layer and, importantly, that they are arranged in a more regular structure i.e. lying anterior to posterior with respect to the whole vocal fold. This strong ordering of the fibres in addition to the inherent strength of the elastic fibres themselves means that this layer is largely responsible for the mechanical properties of the whole vocal fold. Fields and Dunn (1973) calculated that about three times less stress is required to produce a measured length increase in these elastic fibres as compared to collagen fibres.

- Lamina Propria (Deep Layer). In common with the above layer, the fibres in the deep layer are arranged parallel to the edge of the vocal fold but they differ in their composition. They are mostly collagen fibres and therefore combine a high level of flexibility with low elasticity. Hirano (1981) likens this layer to a bundle of cotton thread.

- The vocalis muscle. The largest part of the vocal fold, by volume, is made up of part of the thyro-arytenoid muscle (the vocalis). Despite some controversial findings in early studies, (Goerttler, 1950; etc.) the fibres of the vocalis are now believed to run parallel with those of the intermediate and deep layers of the Lamina propria.

1.1.3 Vocal fold growth rate summary

The vocal folds of the adult female have a smaller cross-section than those of the adult male because of a lesser amount of tissue, and the angle between the thyroid laminac (thc alae) of adult females is significantly greater than that of adult males (approximately 120° compared to 90°). At puberty in both sexes, but more so in males, the walls of the larynx become reinforced, the laryngeal prominence (Adam's Apple) becomes more obvious and the vocal folds grow thicker and longer. Kaplan (1971) has outlined the developmental growth trend of the vocal folds; at birth the length of the vocal fold is said to be around 3 mm, within two months this has increased to 5 mm. By the age of five years the vocal folds are still only around 7.5 mm in length and by the onset of puberty the figure is around 9.5 mm. The vocal mutation which occurs at puberty and involves a drop in average fundamental frequency of one octave in males and two tones in females appears to coincide with the period at which males and females start to diverge markedly in the length of their vocal cords, on average, at some point between 10 and 14 years. In adults, the vocal fold length is 17 to 21mm in males, and 11 to 15 mm in females.

Hirano (1983) concurs with the figures for the size of the mature adult larynges but cites linear growth rates of vocal fold membranous length of 0.7 mm per year for males and 0.4 mm per year for females up to the age of 20 years. The membranous length (L_m) of the vocal folds is that part of the folds where phonatory vibration occurs and differences in L_m have been shown by Titze (1989) to be related to fundamental frequency variation. A scaling factor of 1.6 based on membranous length accounts almost entirely for differences in mean fundamental frequency, aerodynamic power and mean airflow between adult males and females. This scaling factor corresponds well to Hirano's ranges of male and female vocal fold lengths shown above. The difference in cartilaginous length (L_c) of the vocal folds gives rise to another, smaller scaling factor of 1.2 which also relates to overall larynx size. L_c, in combination with L_m, is claimed to account for glottal efficiency and amplitude of vibration (Titze, 1989). The question of whether the growth rate of prepubertal larynges is truly linear or shows developmental differences between the sexes will be further addressed in section 1.2.1.

1.1.4 Adult fundamental frequency

Kent and Read (1992: 154) state that, "women's voices are on average about one octave, or about 1.7 times, higher than mens'". This statement is potentially confusing sice a pitch difference of one octave represents an exact doubling of fundamental frequency. It is perhaps the case that the scaling figure of 1.7 is the more accurate of the two descriptions and the convenient measure of the musical octave has been included as a perceptual aid. Eguchi and Hirsh (1969) report mean fundamental frequencies of vowels excised from sentences for adult males and females of 124.2 Hz and 220.9 Hz respectively. The ratio between these values (1:1.78) is very close to Kent and Read's scaling figure of 1.7.

It is not possible to accurately characterise the typical Fo of adult males or females by means of a single average figure or even by means of a single pitch range because not only is there a considerable amount of Fo variation between speakers, but there are also large differences in Fo amongst different types of speech sample. For example, for both male and female speakers, the production of isolated vowels involves a significantly higher mean fundamental frequency than spontaneous speech, with the figure for oral reading falling between the two. Furthermore, vowel height has been shown to influence the intrinsic fundamental frequency of isolated vowels. (Peterson and Barney, 1952; Lehiste and Peterson, 1961; Sorenson, 1989)

Ohala (1983) has made some interesting comments pertaining to the status of pitch-use as a language-universal feature which serves to enhance the existing biological differences between the sexes. He points out the tendency for high pitch events to correspond with linguistic purposes which are characteristic of stereotypical 'female speech': the use of high pitch often relates to politeness whereas lower pitch is associated with assertiveness; high (or rising) intonation patterns tend to accompany questions whilst low (or falling) intonation patterns tend to be linked with statements. In tone languages high tones usually represent small objects and low tones carry an implication of largeness. Ohala continues by quoting examples of comparative animal development and suggests that because the sexual dimorphism evident in the vocal tract growth occurs simultaneously with the onset of other secondary sexual characteristics under hormonal control at puberty, the evolutionary path followed by the vocal tract may be said to have improved the ability of males, as hunters, fighters and general protectors of the family, by giving them voices with lower pitch which suggest to the listener greater apparent body size. Other non-linguistic secondary sexual characteristics function in the same way (e.g. the development of facial and body hair in adult human males may be analogous to the growth of certain feathers by male birds i.e. to enhance the appearance and stature of the male over the female).

Some of Ohala's comments may not rest comfortably in today's climate of political correctness, however we can extract the important point that there is evidence which seems to strengthen the link between the perception of pitch and speaker sex.

Woods (1992) investigated the relationship between speaker sex and certain intonational features of language. She found that females (both adults and children) used significantly more rising and high-fall tones than males and that males used more level tones than females. Traditional approaches to the study of intonational function often attempt to explain sex-linked variation in terms of grammatical or discourse-based facts. Specifically, the use of falling or level tones are thought to be appropriate for declarative constructions and rising tones tend to characterise interrogative structures. A number of researchers (e.g. Lakoff, 1975) have therefore suggested that the greater use of rises by females may be due to the possibility that they ask more questions than males - the variation could thus be seen to be fundamentally of a syntactic rather than tonal nature. The results of Woods' (1992) study, however, showed that women did not use more interrogative forms than men, in fact the reverse was true, men were seen to ask more questions than women. Furthermore, 48% of the polar interrogative forms which occurred in the study were not accompanied by a rising tone. Woods concludes that the preponderance of rising tones in the speech of females could not be attributed to their alleged preference for interrogative syntax but instead suggests that the intonational feature of tone is socially indexical of speaker-sex.

Whereas average pitch is considered to be a reflection of the rate of vocal fold vibration and therefore is influenced by anatomical and physiological factors, (for example, length, thickness and tension of the vocal fold), pitch range cannot be clearly explained by means of an appeal to anatomy. When we consider that speakers of different languages may adopt very different pitch patterns in everyday speech then we may have to look further than simple physical size or shape differences to account for pitch range variation between the sexes. Woods (1992: 93) found a clear difference in pitch ranges of adult men and women and noted,

> "...if, as previous research suggests, pitch of voice is language specific, and if it is possible to change one's pitch of voice at will (e.g. vocal remodelling in transsexuals), then it seems highly likely that pitch is conditioned by more than just larynx physiology."
>
> (My parentheses)

1.1.5 Adult formant frequencies

> "The vowel formants represent the acoustical resonant properties of the vocal tract as shaped in articulation by the tongue."
> (Eguchi and Hirsh, 1969).

Like Fo values, it has been shown that adult men exhibit significantly lower mean values for formant frequencies than adult women (Peterson and Barney, 1952; Ladefoged and Broadbent, 1957). Whilst differences in laryngeal fundamental frequency between males and females can be ascribed largely to differences in the dimensions of the vocal folds, vowel formants are vocal tract resonances whose central frequencies depend on features of the whole vocal tract, in particular, the size and shape of the vocal cavity. These features are determined, in part, by the individual head and throat size of the speaker and, as females tend to be smaller than males, their consequently smaller vocal tracts will give rise, on average, to higher formant frequencies (Coleman, 1971).

Mathematical prediction of formant differences based on the observed anatomical evidence is not straightforward because, as Fant (1966) pointed out, there are differences other than size between the vocal tract anatomies of men and women. Fant (1973) went on to show that although there are differences in the relative dimensions both of the pharyngeal and oral regions of the vocal tract, the sexual differentiation is greater for the pharynx. Using articulatory data from Russian he reported that the pharynx of females is approximately 2.3 cm shorter than the male measurement but the mouth cavity is only 1.3 cm shorter. This being so, we would expect, therefore, that male and female formant frequencies could not be related by a single scaling factor as the production of different vowels would involve the constriction of different parts of the vocal tract with different acoustic results across the differing anatomies of males and females. Indeed, the sexual differences have been shown to be both vowel- and formant-dependent. Fant related the frequencies of the first formants of males and females by means of a scaling factor (k) according to the following formula:

$$k_1 = \left[\frac{\text{F1 of female}}{\text{F1 of male}} - 1\right] \times 100$$

Other formants could be similarly compared by substituting the appropriate values for F2, F3, F4 etc. As anticipated, Fant determined that k varied as a function of vowel identity, with the average scale factor being about 18%. Importantly, Fant linked some of the anatomical differences in vocal tract structure between men and women to acoustic differences as represented by the pattern of k. He showed that the values of k_1 and k_2 (the scaling factors for F1 and F2 respectively) were small for rounded back vowels, that k_1 was small for articulatorily close or highly-rounded vowels and that k_1 was large for very open vowels. He accounted for these findings neatly by implicating the greater relative length of the male pharynx than the female (Fant, 1973; 1975). Further experimentation by Bennett (1981) repeated Fant's relative k-factor findings with 7 - 8 year old children however Kent and Forner (1979) found that whilst k_2 and k_3 were equivalent to Fant's results, the vowels of the children in their experiment displayed k-factors for F1 which were in reverse order with respect to Fant's k-factors.

As is implied by the quotation at the start of this section, the formant frequencies of vowels are determined by the configuration of the vocal tract which, in turn, depends on two factors - the organic make-up and the precise articulations of the speaker. It is possible to consciously modify one's vocal tract (i.e. articulate) in such a way as to "artificially" raise or lower one's formant frequencies. Mattingly (1966) demonstrated that the observed acoustic differences between men and women could not be wholly accounted for by the average anatomical vocal tract differences. He correlated the distributions of the formants of a number of vowels and found a far lower set of correlations than would be expected on the basis simply of the influence of vocal tract size. It seems that "adult men and women may modify their articulators, lowering or raising their formant frequencies, to produce voices that aim toward male-female archetypes." (Sachs, 1975: 154). For example, Sachs, Lieberman and Erickson (1973) describe how lip-rounding will effectively lengthen the vocal tract and result in lower formant frequencies whilst lip-spreading will shorten the tract and result in higher formants. They point out that "the characteristic way some women have of talking and smiling at the same time would have just this [latter] effect" (Sachs, Lieberman and Erickson, 1973: 81).

These voluntary (although perhaps subconscious) adjustments to the vocal organs in order to bring about formant frequencies which are stereotypically male-like or female-like can therefore be seen to be analogous to the pitch range phenomenon referred to in the previous section. Females tend to speak in a way which identifies them, by means of pitch, pitch range and formant frequencies, as being women. Some of these contributing factors have their roots in genuine anatomical dimorphism between the sexes, others are more directly manipulated by individuals in order to simulate naturally occuring sex-differences or to enhance the difference between the sexes.

1.1.6 Other parameters contributing to the male / female vocal difference

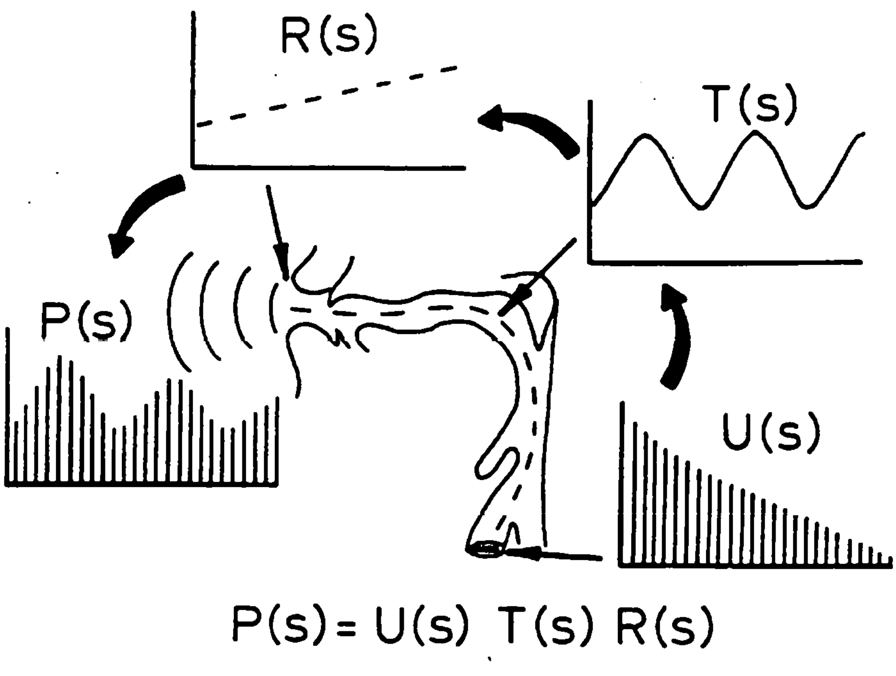
Despite the obvious importance of the contributions of the fundamental and formant frequencies to the characterisation of male / female voice differences and despite the relatively large body of work that has been carried out in these fields, various investigators have identified other acoustic / physiological parameters which have a part to play in the issue (Klatt and Klatt, 1990; Childers and Wu, 1991; Wu and Childers, 1991).

The role played by formant bandwidth and formant amplitude and overall spectral shape has not been widely investigated. Childers and Wu (1991) found vowels spoken by adult males to have significantly narrower formant bandwidths and higher formant amplitudes than vowels spoken by females. They related these findings to the greater degree of spectral tilt[1] in female speakers. No fully comprehensive studies into the detailed acoustic parameters of prepubertal children's speech have, as yet, been published.

1.1.7 Relative contributions of source versus filter components to the vocal sex-difference

According to the linear source-filter theory of sound production, the periodic modulations of the transglottal airflow (or volume velocity) which are associated with the opening and closing movements of the vibrating vocal folds represent the sound source - this is represented by U(s) below (Fant, 1960). This source feeds into the supraglottal vocal tract which acts as an acoustic filter and alters the output characteristics of the source. The precise output depends upon the specific configuration of articulators which defines the resonance properties of the filter (the transfer function) - this is represented by T(s) below. U(s) and T(s) combine with the radiation characteristics, R(s), of the vocal tract to yield the output spectrum, P(s). Figure 1.1 describes the source-filter interaction in the production of a vowel.

Figure 1.1 Source Filter theory of vowel production (from Kent and Read, 1992)



[1] See section 1.1.7 for an explanation of spectral tilt

By the time that a sample of speech arrives at our ears, therefore, the source and filter have interacted to produce a composite waveform. This waveform has a spectrum which contains peaks and troughs rather than the smoothly declining source spectrum, and it is largely this variation in energy distribution across the spectrum which results in our perception of different vowel qualities.

When presented with a vowel, the auditory centre in our brain perceives only a single quality, however, it is possible to experimentally deconvolute the composite acoustic output into source and filter components. The process of inverse filtering is essentially a reversal of the source-filter model of speech production as described above. The convoluted (i.e. source + filter) acoustic output is applied to a filter whose resonance characteristics (transfer function) is the exact inverse of the vocal tract transfer function. Assuming the correct specification of complex poles and zeros, this technique should theoretically yield the simple voice source spectrum as the effects of the inverse filter and vocal tract filter should cancel each other out (Gobl, 1988).

Both parts of the source-filter model are candidates for displaying sex-differences in adults. The filter component represents the vocal tract transfer function which in turn is defined by the relative size, shape and position of the supraglottal vocal organs. As the vocal tracts of adult males are generally larger than those of females following puberty, the resonant frequencies of the transfer function are lower in males. As we have seen, this is seen in the lower formant frequencies of adults males.

On the other hand, any sex-difference in the source component will represent a corresponding difference in the mode of vocal fold vibration and general shape of the glottal pulse. With the exception of Fo variation, there has been considerably less experimental attention given to glottal source characteristics than to properties of the vocal tract filter (Ní Chasaide and Gobl, in press; Ananthapadmanabha, 1984; Ladefoged, Maddieson, Jackson and Huffman, 1987). Spectral analysis of adults' vowels indicates that the waveform of the volume velocity through the glottis during a cycle of phonation is prone to considerable inter-subject variation and may be an important factor affecting speech quality (Holmberg, Hillman and Perkell, 1988;

Childers and Wu, 1991). The glottal excitation wave of adult males is typically asymmetrical with the closing phase representing only 20-40% of the total pitch period and it frequently shows an obvious bulge in the opening phase of the wave (due to source-tract interaction), whereas the female waveform tends not to show any asymmetry indicating less, or no, source-tract interaction (Fant, 1979).

One speech scientist who conducted a number of studies which aimed to assess whether the source or filter component of the voice is the most important carrier of speaker sex information is Martin Schwartz. He carried out a number of studies involving both vowels and fricatives with varying degrees of voicing and voicelessness. In his first experiment (Schwartz, 1968) he played recordings of adult males and females speaking the voiceless fricatives /f/, /θ/, /s/ and /ʃ/ to listeners who made judgements of the gender of the speakers. The rates of correct identification for the fricatives were 93% and 90% for /s/ and /ʃ/ respectively, however only 74% and 69% for /f/ and /θ/. Schwartz explained this significant difference in recognition rate by pointing out that /s/ and /ʃ/, being sibilant fricatives, displayed relatively strong spectral peaks at formant-like frequencies whereas no such peaks are present in the flat, broad-band spectra of /f/ and /θ/ which are non-sibilant. Clearly Schwartz's explanation lends support to the belief that vocal tract resonances play the strongest part in gender identification from voice as firstly, all of the fricatives were reasonably well identified despite a complete lack of fundamental frequency information; and secondly, the highest identification rates were seen in those sounds where "there was a greater proportion of the vocal tract in front of the noise source and, therefore, more information available about vocal tract characteristics." (Temple, 1988; [15]).

In a similar experiment Ingemann (1968) discovered high rates of correct gender identification using /h/ as a stimulus. This finding is consistent with the rationale proposed by Schwartz in that the place of articulation of /h/ is the glottis and so the sound is effectively a voiceless vowel which coarticulates with the following voiced vowel. It is not surprising therefore that the rate of correct identification (and the male - female acoustic difference) of /h/ approaches that of regular voiced vowels.

In Schwartz's second relevant study (Schwartz and Rine, 1968) he turned his attention fully onto vowels. In listening tests judges identified speaker gender from the whispered vowel /i/ in all of the eighty stimuli and identified it in 76 out of 80 cases for /a/. The results of these two experiments by Schwartz were a major piece of evidence in favour of vocal tract information in the debate as to which of the source or filter components could be said to be the most useful predictor of gender.

Coleman (1971) showed that listeners could reliably recognise the sex of a speaker who spoke the vowels /i/ and /u/ and used an electrolarynx[2]. The implication of this finding is that listeners are able to correctly identify gender even when the fundamental frequency was held at the same level for all speakers (due to the electrolarynx).

In a further experiment, however, Coleman attempted to reduce the amount of prosodic information available by presenting his stimuli backwards. Using correlational analyses he was able to determine that,

> "...listeners were basing their judgements of the degree of maleness or femaleness in the voice on the *frequency of the laryngeal fundamental.*"
>
> Coleman (1973; [6] - my italics)

Several years later, Norman Lass led a team of researchers in an experiment which took Schwartz's original ideas one step further. In this study listeners had three sets of stimuli to judge: as in Schwartz's study there were samples of isolated voiced and whispered vowels, however in addition listeners were asked to identify samples of voiced vowels which had been low-pass filtered at 225 Hz to eliminate higher harmonic patterns and hence remove formant frequency information. Interestingly, results showed that the voiced and filtered stimuli had significantly higher rates of correct identification than the voiceless stimuli and Lass et al. were forced to conclude that,

[2] The electrolarynx is a manually operated electrical buzzer which, when held to the neck, can simulate the vibrations of the larynx. It is used, for example, by laryngectomee patients who have had some or all of their laryngeal organs surgically removed following pathology or localised trauma.

> "the laryngeal fundamental appears to be a more important acoustic cue for speaker sex identification than the speaker's resonance characteristics (formants)".
>
> (Lass, Hughes, Bowyer, Waters and Bourne, 1976; [678])

Thus we have seen that there appears to be evidence to support the status of both source (fundamental) and filter (formants) as the more important influence in the process of gender identification. Perhaps the truth of the matter is that Fo and formant information *combine* to provide the listener with relevant acoustic information regarding the speaker's gender. Childers and Wu (1991) hypothesised that Fo may be the major cue to gender but if the Fo information should be surpressed (for example, due to a vocal deviation) then vocal tract resonance characteristics may surpass Fo in the recognition process. Following extensive mathematical modelling of the perceptual gender identification strategy, they concluded that both fundamental frequency and formant characteristics proved to be robust cues to speaker sex but that there was "considerable redundant information concerning gender [...] embedded in the formant and fundamental frequency features." They found that the highest identification rate (98.1%) was extracted when using the second formant frequency as a single feature. When considering grouped features, the third formant group (frequency, amplitude and bandwidth of F3) yielded a perfect identification rate of 100%. Both of these figures surpass the best rate of identification due to Fo information (96.2%), however it seems clear that the differences reported in this study were very small and that although Fo might not statistically be the single best acoustic parameter for identifying speaker sex it still contains more than enough information to allow correct gender identification.

In summary, a number of experiments have been reported which have shown that, depending on their focus, fundamental frequency or formant information are highly influential in the listeners ability to correctly assign gender to an adult speaker. It is not possible at this time to firmly name one or other of the acoustic parameters as the primary influence and, moreover, it may be the case that under normal conditions both factors are used in combination whilst individually retaining the power to carry sufficient information for gender recognition should the need arise.

1.1.7.1 Contribution of spectral tilt / breathiness to perceived sex-differences

The parameter of spectral tilt describes the roll-off effect of the envelope of a glottal source spectrum. In an ideal modal phonation, the amplitude of the harmonics declines at a rate of 12 dB per octave, i.e. for every doubling of frequency, the amplitude decreases by 12 dB (see figure 1.1). Strictly speaking, the result of inverse filtering is not equivalent to the true glottal flow because the effect of lip radiation has not yet been accounted for. The consequence of lip radiation with respect to the voice spectrum is a relative amplification of higher frequencies by 6 dB per octave. If this contribution is not neutralised, the result of the inverse filter will correspond to the differentiated glottal flow (glottal flow derivative). The solution, therefore, is a simple process of mathematical integration. With the lip radiation accounted for, the resulting spectral slope will much more closely approximate the ideal value of -12 dB / octave. The slope of the glottal flow derivative is thus around -6 dB / octave.

Holmberg, Hillman and Perkell, (1988) discovered a number of statistically significant differences between male and female glottal waveforms which, in summary, suggest a steeper spectral slope for females - an acoustic state which is often represented perceptually by an increase in breathiness. The voices of women are often classed as being significantly more breathy than those of men. Amongst the acoustic / physiological correlates of vocal breathiness are :

1) increased transglottal leakage (i.e. the escape of air through the glottis even during the closed phase[3] of the glottal cycle

2) a relatively more dominant fundamental frequency (first harmonic) and therefore steeper spectral tilt.

3) a larger open quotient[2]

4) an increased level of noise in the frequency regions between the formants.

---

[3] See appendix 1 for an explanation of the terms "closed phase" and "open quotient"

These factors and others cited by Kent and Read (1992) are all interconnected to some extent. The fact that there is a larger open quotient in breathy vowels means that there will be greater airflow through the glottis. This may amount to a 60% increase in average airflow over modal vowels. In turn, this increased transglottal leakage produces acoustic turbulence at the glottis, and elsewhere in the vocal tract, and results in aperiodic noise interspersed among the harmonics of the voicing spectrum.

It was calculated by Bless, Biever and Shaikh (1986) that women were up to four times more likely to display a posterior glottal chink during the closed phase of the glottal cycle than men. Klatt and Klatt (1990) used this vocal setting in the female to account for an increased bandwidth in the first formant which, according to the authors, was so pronounced as to be "sometimes obliterating the spectral peak at F1 entirely" (Klatt and Klatt, 1990; [835]). The same authors also comment on the acoustic coupling of the subglottal region in females and cite this as a problem when analysing males and females together, however, there is little evidence that the tracheal coupling will significantly influence the spectral output given a normal closed quotient.

The relative dominance of the first harmonic (fundamental frequency) over the second harmonic in women is believed to be due to a difference in phonatory mode between the sexes. In a normal (non-breathy) vowel the vocal folds close with such speed and momentum that the transglottal passage of air is sharply bisected. The glottal pulse which results from this abrupt vocal fold closure has higher harmonics which are rich in acoustic energy. In a breathy vowel, on the other hand, the weaker adductory motion of the vocal folds give rise to a glottal pulse which is not characterised by either a sharp onset or a sharp closure. A pulse which lacks the important, marked cessation of the airflow will be relatively weak in acoustic energy in the higher harmonics and relatively strong in energy of the fundamental component of voicing.

Ladefoged and Antoñanzas-Barroso (1985) compared two spectral measures of breathiness with a waveform measure in speakers of !Xóõ (which has breathiness as a phonemic contrast in its vowels)[4] and correlated the results with perceptual judgements of breathiness. They found that both of their spectral measures (H1-H2 and H1-F1) distinguished breathy from modal vowels with the breathy vowels displaying higher values for each measure. It was shown that the distinction between the two voice qualities did not correlate with absolute differences in H1-F1 or H1-H2, but relied rather on a relative scaling effect - what was modal for one speaker was considered breathy for another. The H1-F1 measure correlated very highly with perceptual ratings of breathiness (r=0.93) indicating that listeners were making use of the increased spectral tilt in the breathy vowels to distinguish them from modal phonation.

The authors also measured breathiness from the speech waveform by determining the contribution of the aperiodic, turbulent airflow component relative to the regular, stable vibrations of the vocal folds. This measure also separated breathy from modal vowels to a satisfactory degree, however when correlated with listeners' perceptions, it was found to be a poor predictor of breathiness (r=0.57). It seems that for American listeners attending to speakers of !Xóõ at any rate, the manner of vocal fold vibration as reflected in measures of spectral slope is more important than the amount of aperiodicity in the voice in distinguishing breathy from modal phonation.

In adult males the vocal folds are heavier and there is a certain amount of vertical displacement during phonatory vibration with respect to females. This sex-difference permits the males to achieve a stronger and more complete vocal fold closure which is slightly out of phase and has a relatively longer closure period. The resulting waveform is thus asymmetrical and displays enhanced harmonics above the fundamental. However, the increased glottal leakage in women's speech, interestingly, may not be the direct consequence of an anatomical difference - as we have seen, there is some physiological evidence to support the contention that women may not maintain as tight a glottal closure as men, however, as females speak with a higher fundamental frequency than males, they are not under the same "pressure" to generate strong higher harmonics to make themselves heard. The glottal relaxation of the female (relative to the male situation) probably develops as a

[4] !Xóõ belongs to the Southern Khoisan family of languages and is principally spoken in Botswana

socio-cultural cue which is seen as a socially desirable vocal attribute in adult females.

The nature of this proposed phonatory preference in females was investigated by Henton and Bladon (1985). They pointed out that the phenomenon of breathiness is widespread in the world's languages and is used in many cases to form the basis for distinctive linguistic contrasts. Breathy phonation is also a marker of certain laryngeal pathologies and is a strong perceptual cue for identifying certain speech disorders. However, normal speakers of British English, which does not involve any phonemic contrasts between different voice qualities, have no strong organic or linguistic basis for using breathy phonation. Henton and Bladon (*ibid*) cite a number of reasons why it is unlikely that a speaker would favour the use of breathy over normal voicing: breathiness often enforces reduced vocal intensity and lower pitch which can give rise to greater monotonicity in speech; also, because breathy phonation requires a relatively low level of longitudinal tension in the vocal folds, high-pitched breathy vowels are less likely.

Breathiness is also detrimental to accurate speech perception. It forms a background of noise across the voice spectrum (lowering the signal-to-noise ratio), and the increase in first harmonic intensity will interact with the low first formant of a close vowel adversely affecting its intelligibility.

The authors formulate the paradox succinctly, "British women are apparently adopting articulatory postures which not only undermine the perceptibility of their speech, but also tend to make it less varied and more monotonous." (Henton and Bladon, *ibid*; 225)

They go on to suggest a hypothesis to account for the problem of why women should voluntarily choose to adopt a type of vocal fold vibration (i.e. breathy) which was significantly less efficient than that of modal voice. In our culture, a breathy voice is perceived as 'sexy' and this, in turn, is linked with arousal. There is a suggestion that the connection between breathiness and arousal may be a physiological one in that when sex hormones are released from the brain, a number of other secretions take place around the body, including at the larynx. An increase in the lubrication of the vocal folds (specifically the epithelium and lamina propria - see section 1.1.2.3) would limit their full closure and give rise to a breathy voice.

It is not suggested that adult females are inherently more aroused than males, rather that they choose to imitate the voice quality which is associated with arousal. Sociological theory would predict that a woman who speaks with a voice that sounds as though she is sexually aroused will tend to be treated with greater attention and meet with greater approval by males than if she were to use modal voice. Seen in this light, breathiness can be regarded as forming part of the courtship ritual, "A breathy woman [uses] her paralinguistic tools to maximise the chances of her achieving her goals, linguistic or otherwise." (Henton and Bladon, *ibid*; [226])

1.2 GENDER DIFFERENCES IN THE VOICES OF PREPUBERTAL CHILDREN

1.2.1 Anatomical and Physiological development of the vocal folds in prepubertal children

The major period of anatomical development of the vocal mechanism is during the years from birth to post-puberty and, despite some disparate findings in the literature, the consensus seems to be that there are no significant structural differences between the vocal organs of males and females prior to puberty.

At birth the infant larynx is smaller in real terms than the adult counterpart although, in proportion to the rest of the body, it is larger. It is also softer and more flexible at this stage. Kaplan (1971; [244]) describes the movement of the larynx with age, "the larynx descends and its lumen is displaced posteriorly from the vertical." He points out that it is the growth of the back of the tongue which contributes mostly to this movement by pushing the inlet to the larynx backwards.

Kaplan cites the period from birth to three years as the major interval of laryngeal growth after which there is little further development until puberty. Again, he claims that there are no major differences in larynx size between the sexes before twelve years of age.

The vocal folds develop at a slightly different rate from the larynx as a whole. At birth the folds are obviously shorter than the adult's folds, both in real terms and relatively but, unlike the larynx, they grow at a roughly linear rate, approximately until the age of 20, which Hirano (1983) has estimated to be 0.4 mm per year for the female and 0.7 mm per year for the male. The vocal folds are around 3 mm in length in new-borns and, according to Kaplan (1971), grow to 5 mm within the first two months of life and to 7.5 mm by the age of 5 years. These data do not correspond with Hirano's calculation of growth rate and seem to ally themselves more to a quantal theory of growth. The quantal theory holds that most of the growth in vocal fold length occurs at puberty, particularly in the male. The rapid drop in fundamental frequency associated with the vocal mutation and the relatively sudden prominence of the Adam's apple are evidence for rapid laryngeal development in the pubertal male. Titze (1989) cites measurements made by Kahane (1978) of increase in overall larynx length which amount to 62% from age 10 to 16 years in males and 34% from age 12 to 16 in females. In comparison, Titze also cites Hirano's data of linear growth of vocal fold membranous length as 48% for males and 27% for females of the same ages.

Titze accounts for these differences in terms of variations in measurement methodology. The evidence for or against either the linear or quantal theories is not yet conclusive, however as I shall show, the quantal theory best seems to fit the available measurements.

Figures 1.2, 1.3 and 1.4 show the vocal fold growth rates quoted by Hirano (linear) and Kaplan (quantal) plotted as line charts. Figure 1.2 shows the growth rates in males. The linear function can be seen to increase smoothly from a value of 3mm at birth to an averaged adult value of 17mm at a constant rate of 0.7mm per year. In contrast, the quantal function increases rapidly during the first three years and then begins to level off during the childhood years. At some time around the age of 15 years there is a second major growth spurt which results in a doubling in the length of the vocal folds within a short period of time (within one year according to Sinclair (1985)). This is the adolescent growth spurt which gives rise to the vocal mutation.

Figure 1.3 shows the growth rates in females. Once again the linear function is a simple straight line, this time with a growth rate of 0.4mm per year. The quantal function for females, as with males, shows extremely fast growth in the early years of life. Following this period the growth rate declines, although less rapidly than the male rate, until the linear and quantal curves converge at age 15. From this stage, the increase in vocal fold length mirrors the linear growth rate of 0.4mm per year. The two theories therefore agree that the developing female vocal fold reaches a level which appears to be around 85% of the mature vocal fold length by the onset of puberty. Figure 1.4 is a magnification of the left-hand half of figure 1.3. It shows, in greater detail, the curves described by the linear and quantal theories for the first eight years of life in females.

Figure 1.2 Quantal and Linear rates of vocal fold growth in males

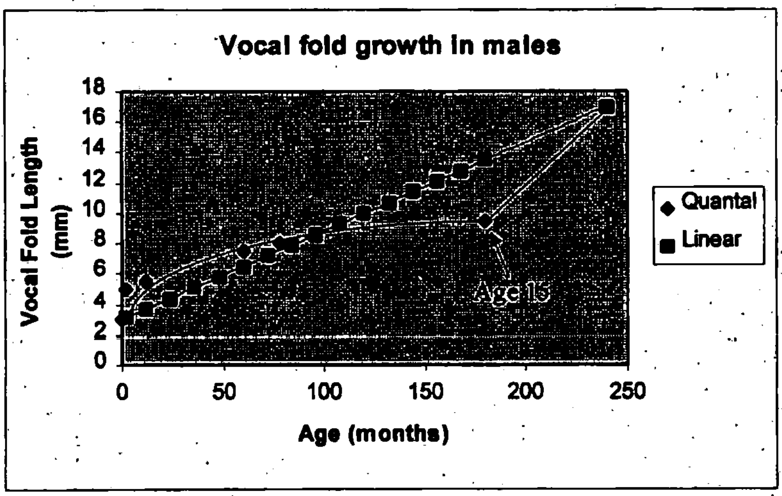


Figure 1.3 Quantal and Linear rates of vocal fold growth in females

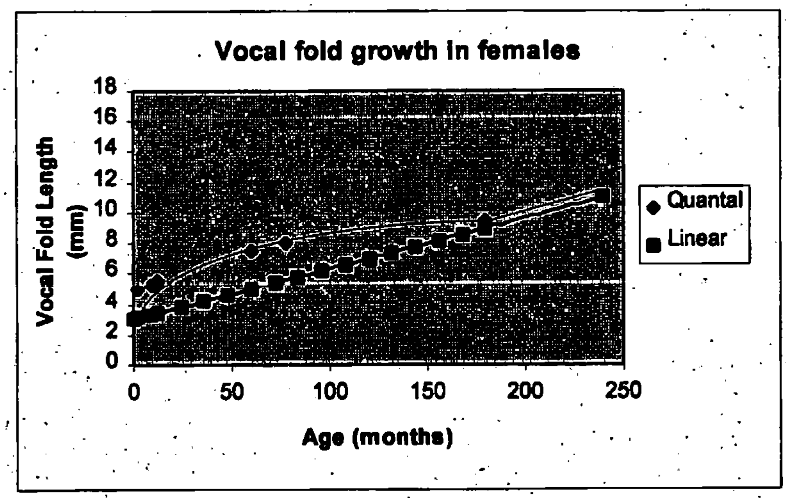
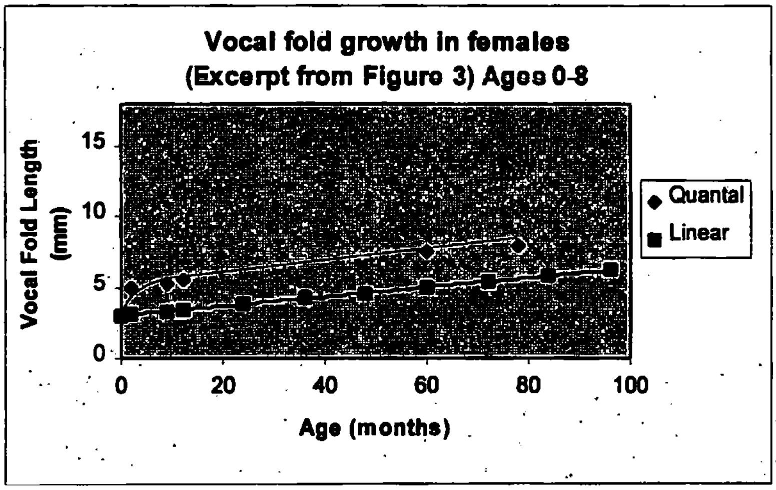


Figure 1.4 Magnification of section of figure 1.3 (Quantal and Linear rates of vocal fold growth in females)



The rapid vocal fold development in the first eighteen months after birth (which is a consequence of the high general level of anatomical and physiological maturation during this period) is more clearly seen. From around 20 months, the growth rate becomes nearly linear although with a slightly steeper gradient than that suggested by Hirano.

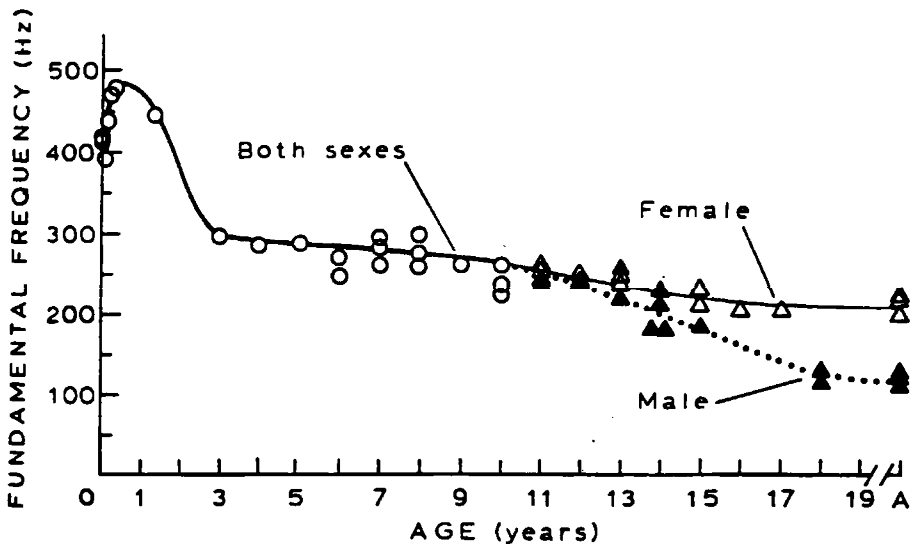
Given an acoustic theory of sound production which inversely relates vocal fold length and mean fundamental frequency, (e.g. Titze, 1989) it is reasonable to attempt to model the course of development of a hypothetical child's average Fo from the different growth rates which we have been considering.

Intuitively, the linear theory appears to be an oversimplification of the developmental routes of boys and girls. It would predict that there is always a sex-difference in vocal fold length, and by implication fundamental frequency, and that it increases linearly with age. Within this framework the pattern of fundamental frequency should simply decline proportionately as age increases.

On the other hand, the quantal theory would predict that there would be a sharp drop in fundamental frequency in the first 2 or 3 years of life. Following this period the average Fo levels should more or less stabilise until puberty at which point there should be another rapid decline to the adult values.

Kent (1976) has collected the results of a number of independent studies which investigated fundamental frequency development. Figure 1.5 shows his Fo findings plotted as a function of age for boys and girls.

Figure 1.5 Mean fundamental frequency plotted as a function of age of subject. The data for males and females are combined until age 11. Original figure from Kent (1976).



It is clear at a glance that this pattern of development corresponds much better to the quantal theory of vocal fold growth than to the linear theory. There is a small rise in Fo during the first four months of life which may be due to the fact that it is during these first few months of life that the infant is attempting to gain sufficient control of his or her laryngeal mechanism to cause intonational effects. By the end of the first year, a child is said to be able to produce at least seven distinct intonational forms (Tonkova-Yampol 'skaya, 1973). After this brief period of large-scale Fo variations, the pattern settles to the course predicted by the quantal theory. From the ages of about 1 to 3, fundamental frequency drops quite sharply as the vocal folds grow quickly. At the age of 3 the rate of Fo decline slows markedly and from this age until the onset of puberty, the slope is close to linear (as in figures 1.2, 1.3, 1.4).

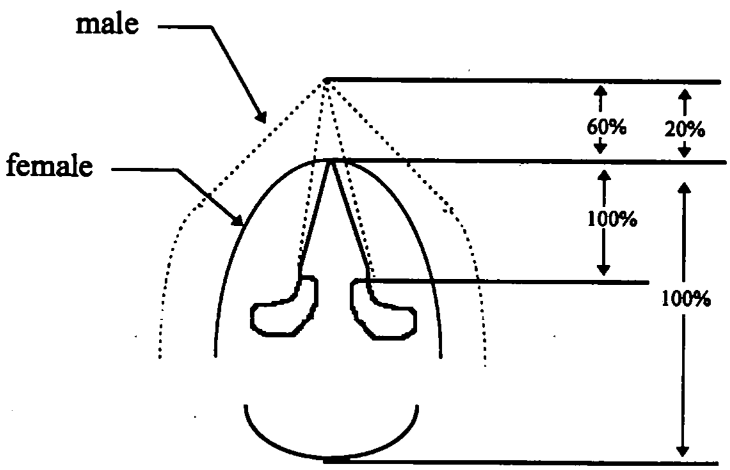
Whereas Kaplan (1971) describes the first three years and puberty as the main periods of laryngeal growth and Kent's Fo measurements seem to back this up, Negus (1962) claims that the ages of 3-5 years serve as one of the primary growth terms. These apparently conflicting positions may be reconciled by Titze (1989;

[1701]) who pointed out that "the age of 3 is the median age at which the vocal ligament begins to develop". It is possible that the increase in vocal fold stiffness which accompanies ligamental growth may cancel out any effect of a length increase on the mean Fo. In other words, Negus may be right insofar as there is major laryngeal growth occurring in the vocal ligament (and perhaps vocalis muscle) during this period, however Kent may also be correct in claiming little increase in fundamental frequency over the same period.

Until now, we have been discussing vocal fold lengths without reference to the internal morphology of the folds themselves. In fact, however, the vocal folds can be sub-divided into at least two sections in the horizontal plane The section anterior to the tip of the vocal processes of the arytenoid cartilage is known as the membranous length or L_m. This is the part of the vocal fold where phonatory vibration takes place and, according to Hirano (1983) the average L_m in adult males is 60% greater than the average L_m in adult females (see Fig. 6). This ties in with the linear growth rates proposed by Hirano - there is around a 60% difference between the male growth rate of 0.7mm per year and the female growth rate of 0.4mm per year. Titze (1989) has defined a scaling factor to account for sex-differences in L_m which he calls β with the value of 1.6. The remainder of the vocal fold is known as the (inter-)cartilaginous portion or L_c. The difference in L_c between males and females is much smaller than β and Titze quantifies it as a scale factor of 1.2 which he designates as α. This scaling also accounts for the difference in thyroid cartilage size, i.e. the male cartilage is, on average, 20% larger than the female (Kahane, 1978).

Titze used both α and β to rationalise sex-differences in a number of acoustic and aerodynamic parameters. It seems that the scaling factor of 1.6 based on membranous length (β) accounts for many laryngeal sex-differences including differences in mean fundamental frequency, aerodynamic power and mean airflow between adult males and females. α , in combination with β, is claimed to account for glottal efficiency and amplitude of vibration. Titze concludes "it has been demonstrated that the male larynx has a greater source strength, but this difference is overcome almost entirely by the fact that the female radiates at higher frequencies...The power can be comparable, however because...the female may deemphasise this high-frequency gain by spreading the vocal processes." (Titze, 1989; [1706]). In addition to reducing the effective power output of the female voice, this will also result in a breathier voice quality (see section 1.1.6.1).

Figure 1.6 Horizontal section of idealised male and female vocal folds showing sex-differences in size. (after Kahane, 1978)



1.2.2 Fundamental frequency in prepubertal children

In addition to the clear-cut anatomical size differences between the vocal organs of adult males and females, there are some equally clear-cut acoustic and perceptual differences between aspects of their voices. Whilst the course of development of the organic differences is relatively well-defined, there has been much debate in the literature concerning the developmental processes governing acoustic parameters such as fundamental frequency.

The past thirty years has seen the publication of a number of pieces of research into the problem of characterising the fundamental frequency attributes of preadolescent children (Kaplan, 1971; Murphy, 1964; Luchsinger and Arnold, 1965; Eguchi and Hirsh, 1969; Bennett and Weinberg, 1979a; Sorenson, 1989). Despite this large body of research, the findings are varied and conflicting in their conclusions. Amongst the different studies there is a large selection of differing methodological approaches both in terms of the phonatory tasks being used and the analysis techniques employed. This inconsistency of procedure makes it extremely difficult to draw conclusions concerning sex- or age-related differences from the diverse body of evidence available. Hasek, Singh and Murry (1980) provide a review to that date of the major studies concerning the fundamental frequency of 5-10 year old children. Out of the 24 different studies which they quoted, not one involved the subjects performing more than a single type of phonatory task. This deficiency was the basis for work done by Sorenson (1989) who argued that, as there are major differences in fundamental frequency across different speech tasks in adult speakers, the type of task used by researchers in eliciting corresponding data from preadolescent children may have a considerable influence on the final results. Statistical analysis of Sorenson's results indicate that there is no significant difference between the mean fundamental frequency of prepubertal boys (262 Hz) and girls (281 Hz) when averaged across speech tasks. The Fo values for the different age groups between 6 and 10 years were not statistically significantly different, neither were the Fo values for the different tasks[5]. This latter finding contrasts with the situation in adults where there is a continuum of Fo values with sustained vowel production at the top, spontaneous speech at the bottom and oral reading between the two.

Interestingly, when the data were analysed separately according to the sex of the child, both age- and task-related differences began to appear. Firstly, the average Fo of the boys was found to drop between 6 and 7 years and then to remain stable from 8 to 10 years; no significant change was recorded for girls in the same age ranges. Secondly, Fo values of spontaneous speech were significantly lower than Fo values of either sustained vowels or reading for both boys and girls.

Sorenson's results are open to interpretation both in support of the equivalence of Fo values between boys and girls prior to puberty and, to a lesser extent, in support of an unequal developmental shift in the Fo values of prepubertal males and females. His main findings agree with those studies which report no significant difference in the fundamental frequency of boys and girls between the ages of 6 and 10 years (Eguchi and Hirsh, 1969; Weinberg and Bennett, 1971; Kent, 1976; Bennett and

[5] We should remember, however, that it is possible that a statistically non-significant result may still be perceptually salient.

Weinberg, 1979a) and the mean Fo values are consistent with those stated by Kent (1976) for children of the same age.

The lowering of fundamental frequency in males at the ages of 6 to 7 years seen by Sorenson (1989) suggests a developmental trend in sex-differences which is not without precedents. Bennett (1983), for example, conducted a longitudinal study over a three year period using children ranging in age from 8 to 11 years. She detected age-related changes in mean fundamental frequency at that age range for both boys and girls but these changes did not result in a significant sexual difference in Fo values.

Vuorenkoski (1978) measured the fundamental frequency characteristics of Finnish speakers between the ages of 6 and 40 years, including a number of preadolescent children. Children aged 6, 8 and 10 years were compared and among the boys a significant difference was observed in speaking Fo between the ages of 8 and 10, but no corresponding change for females during the same period was found.

The downward trend in male Fo values seen by Sorenson (1989) supported work carried out by Hasek, Singh and Murry (1980) who found that "a significant difference between the average Fo of preadolescent males and female children begins to emerge by age seven or eight...The male/female difference was directly attributable to a decline in Fo for male children only, beginning around age seven." [1264]. Neither Hasek *et al.*, (1980) nor Sorenson, (1989) offer firm explanations for this relatively unheralded sex-difference, however both parties cite Crelin's (1973) comment which reported sex-differences in larynx size and thyroid angle developing as early as the third year of life. The suggestion is that if these anatomical differences continue to manifest themselves at the same rate, then by the ages of 6, 7 or 8 years they might be large enough to account for the drop in male fundamental frequency. Alternatively or additionally, male children might be prompted to lower their fundamental frequency according to some socio-cultural factor.

The subjects in the experiment performed by Eguchi and Hirsh (1969) were either aged between 3 and 13 years or were adults and were required to produce sustained vowels from which estimates of fundamental frequency were made. Their results indicated, from a value of around 300 Hz at age 3, Fo declines slightly with age, with the largest drop occurring between the ages of 3 and 6 years. It is not possible to draw conclusions regarding sex-differences from this data as all of the male and female subjects are grouped together up to the age of 11, however the lowering of Fo values at an earlier age than those studies previously mentioned is perhaps an indication of the variability of the developmental processes influencing laryngeal production.

Bennett and Weinberg (1979a), in attempting to determine whether differences in average fundamental frequency among preadolescent children in their study could account for their reported recognition performance (see section 1.3), concluded that no significant Fo difference exists between boys and girls. They found instead that despite the listener's ability to discriminate male and female children into discrete perceptual categories, there was a large overlap in the distribution of mean Fo.

Sachs, Lieberman and Erickson, (1973) measured the Fo values of pairs of preadolescent boys and girls matched for height and weight. The rationale for this method was that it was felt that matching two children for height and weight would compensate for organic size differences in the vocal tract, so that as long as both children weighed the same and measured the same for height, irrespective of their age, they would have roughly equivalent anatomies prior to puberty. What they found was that the boys in fact displayed a *higher* mean fundamental frequency than girls (274 Hz compared with 249 Hz). This study will be examined in more detail in section 1.3.

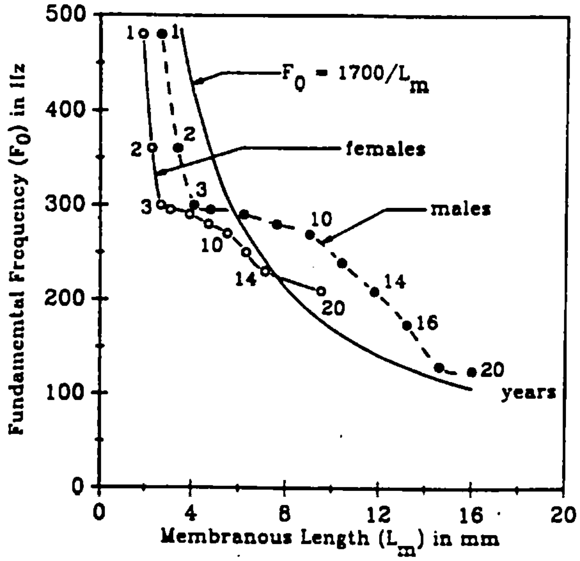
Kent (1976) has investigated the course of Fo development with age and figure 1.5 shows Fo values plotted as a function of age for data which he collated. Clearly, the periods of most extreme Fo movement are: from birth to four months; from 1 to 3 years and the period of puberty (roughly 13-17 years). According to Kent's figures, after falling slightly in the first month of life, Fo rises until about the fourth month and then remains relatively stable until one year of age. The following two years, up to the age of three, is a period of rapid decline in fundamental frequency from above 400 Hz to around 300 Hz, after which time Fo values decrease much more slowly until the onset of puberty and the vocal mutation in males. Kent notes that the periods of rapid Fo movement correspond to marked periods of laryngeal growth as defined by Kaplan (1971).

Titze (1989) highlights an interesting sex-difference by plotting Kent's data concerning fundamental frequency against Hirano's vocal fold membranous length figures (see figure 1.7). In addition to the male and female curves there is a curve representing the theoretical relationship between the fundamental frequency and membranous length (L_m) as would be predicted by a vibrating string model with fixed tension and fixed mass per unit length. This curve derives from the equation

$$F_o = 1700 / L_m .$$

It is interesting to note, from this graph, that before puberty males seem to display the same mean fundamental frequency as females whilst showing consistently larger values of L_m. This is further support that fundamental frequency in children does not vary as a function of vocal fold length until puberty and defends the view of no consistent sexual Fo differences in preadolescent boys and girls.

Figure 1.7 Mean speaking fundamental frequency, Fo, as a function of vocal fold membraneous length, L_m. Solid line without data points is a hyperbola that serves as a first-order "string" model. Original figure from Titze (1989).



The male and female curves in figure 1.7 exhibit a significant 'kick' from their predicted course at around the age of three. This may be due to the onset of vocal ligament growth which Hirano (1983) has estimated to begin at the age of 2 to 4. Titze speculates that this growth of the ligament, combined possibly with a vocalis muscle growth, may help to account for the relative stability of fundamental frequency values between the ages of 3 and 10 years, in that the effects of the increase in vocal fold length during this period might be cancelled out by an increase in overall stiffness of the laryngeal tissue. The continued and rapid decline in Fo associated with the voice break in males is seen quite clearly in figure 1.7.

The feature of pitch variation has not been widely studied with respect to prepubertal children, however Woods (1992) has found evidence of the role of pitch as an indexical marker in two separate functions. She found similar patterns of tonal movement in children as in adults (see section 1.1.4) namely that girls used more rising and high-fall tones than boys and that boys used more level tones than girls. She also found that girls demonstrated wider pitch ranges than boys. Overall, Woods' findings can be interpreted as showing the importance of pitch features to the acquisition of sex-appropriate speech styles in children.

1.2.3 Formant frequencies in prepubertal children

On the basis of experimental investigation, Kirchner (1970) proposed that there is no significant difference in larynx size between prepubertal children who are matched for height and weight. As we have just seen, Hirano (1983) has produced some more recent data which shows small but consistent differences between the vocal fold membranous lengths of males and females in the first ten years of life. Whichever of these studies, if either, is correct, it seems to be the case that there is no clear sexual distinction in fundamental frequency for children during this age range.

On the other hand, the differences in the formant frequencies of adult males and females is largely explained by the fact that males have larger vocal tracts than females. Given the same input, differently sized acoustic tubes (i.e. vocal tracts) will produce different resonant frequencies (i.e. formants). Generally speaking, the larger the tube, the lower the resonant frequency, hence males tend to have lower formant frequencies than females.

Anatomical studies have suggested that mandible length, which constitutes half of the supralaryngeal tract length is equal for pre-pubertal boys and girls (Walker and Kowalski, 1971; Hunter and Garn, 1971) and this has prompted Sachs *et al.* (1973) to state that children who are matched for height and weight are assumed to have the same supralaryngeal vocal tract size and, by implication, the same formant frequencies. This statement appears somewhat extreme given that we can only prove that half of the supralaryngeal vocal tract length (i.e. the mandible) is developing equally in boys and girls. Bennett (1983) made an altogether more reasonable comment when she said that as "mandibular length is highly similar among males and females prior to puberty, the two sexes probably do not differ markedly with respect to *oral cavity size*" [234] (my italics). She goes on to point out that there is evidence of a sexual difference in pharynx size before puberty with boys measuring 2% to 8% longer than girls for one measure of pharyngeal length during the first ten years of life and 13% longer by the age of 16 (King, 1952). It seems likely then that differences in the size of both the oral and pharyngeal cavities combine to set up a relatively large sex difference between the formant characteristics of adults, whereas it is perhaps only the pharynx which contributes to the difference in prepubertal children. If this is the case then we would expect quantitatively smaller formant differences between the sexes in children as compared with adults. This is indeed the situation that Bennett (1983) discovered when she calculated formant scaling factors (*k*- factors)[6] for the 7-8 year old boys and girls in her study. The average *k* - factor across the lowest three formants for five vowels (/i, ɪ, ɛ, æ, ʌ/) was about 10% for the 7 and 8 year old children. In contrast, it was 19% using data from Peterson and Barney's (1952) experiment with adult English speakers and 18% using data from Fant's (1973) experiment with adult Swedish speakers. This finding, that the formant frequencies of prepubertal boys and girls can be compared in the same way but on a smaller scale to adults, is further evidence that formant frequency correlates well with vocal tract size.

In addition to calculating scaling factors between males and females at various age groups, it is possible to calculate a scale factor between adults and children. Thus Kent (1976) used the following formula to calculate the scale factor between the first formant frequencies of adult males and prepubertal children:

$$k_1 = \frac{\text{F1 of children}}{\text{F1 of adult male}} - 1 \times 100$$

When plotted against age, the values of *k* were much greater when using males as the reference value (the denominator in the equation) than females. This is to be expected as it reflects the fact that prepubertal children's voices are acoustically more similar to adult females than to adult males. There is also a greater developmental trend evident in the scaling factors, so that, for example, k_2 (the scaling factor for the second formants) decreases at an average rate of 3.4% for an adult male referent and about 2% for and adult female referent between the ages of 3 to 13 years (Kent, 1976).

Eguchi and Hirsh (1969) obtained a large amount of formant frequency information from prepubertal children but did not separate them by sex until the age of 11. They found that in childhood generally, but particularly between the ages of 3 and 5 years,

6 See section 1.1.5 for an explanation of *k* - factors

the second formant frequency was prone to a greater decline than the first formant frequency with the exception of [a]. Low and middle vowels exhibited more variability in F1 values than high vowels but all vowels showed less variability with increasing age (see table 1.2).

Table 1.2 Mean variability of the first formant frequencies of six vowels at two age groups. (Data from Eguchi and Hirsh, (1969))

Vowel	Mean variability (Hz)	
	At age 3	At age 13
[a]	170	35
[ɛ, æ, ɔ]	100	21-29
[i, u]	55	15

Eguchi and Hirsh (1969) interpret this developmental trend as an indication that articulatory accuracy improves with age. Lindblom (1972), however, disputed the validity of Eguchi and Hirsh's findings as the accuracy of formant estimations made from spectrograms decreases with an increase in the speaker's fundamental frequency. Lindblom found that a hypothetical curve plotting the error (variability) in formant estimation against Fo compared very closely to the curve presented by Eguchi and Hirsh which showed formant variability as a function of age. In other words, as Kent (1976) points out, "measurement error might be a significant factor in the variability data derived from the spectrographic measurements" [434].

Sachs et al. (1973) found that boys exhibited significantly lower formant frequencies than girls for the vowels /i/ and /u/. This study involved a perceptual test to ascertain whether listeners could identify the gender of a child from recordings of his or her voice using children aged from 4 to 12 years. When the best-identified and worst-identified children of each sex were compared, it was found that the most boy-like voices (irrespective of actual biological gender) had higher fundamentals but lower formant frequencies. The correlation between lowness of formants and probability of identifying a boy's voice was significant. Correspondingly, the most girl-like voices had lower fundamentals and higher formant frequencies, although the correlation was not significant in this case.

1.2.4 Other parameters

Commenting on the fact that almost all of the theoretical basis for the current state of technology in the field of experimental speech science was calculated with reference to adult males, Titze (1989) wondered, "if the source-filter theory of speech production would have taken the same course of development if female voices had been the primary model early on." [1699]. Indeed, such is the predominance of the male input to the engineering process that some of the methods used to analyse formants (e.g. Linear Predictive Coding) deal with children's and women's high-pitched voices particularly badly. Despite this situation, there has been very little work done to investigate acoustic / physiological parameters other than formant and fundamental frequencies in prepubertal children.

A methodological quirk of Sachs *et al* (1973) was that Fo and formant frequency measurements were carried out on isolated vowels whereas the gender discrimination was carried out on whole spoken sentences. Therefore, there may be more acoustic/phonetic information present in the sentence than merely Fo and formant frequencies. Indeed, Günzburger, Bresser and Ter Keurs (1987) state that gender recognition "is not based on spectral information only". [52]

Sachs (1975) showed that rates of gender identification are higher when the listener is presented with normal sentences than when presented with either isolated vowels or backward sentences. This being so, sex-dependent cues might be found in the time-domain e.g. such factors as intonation, speech rate, coarticulation.

Further support for this hypothesis is provided by Bennett and Weinberg (1979b) who found that artificially imposed intonational patterns had a significant impact on the perception of gender of children's speech (see section 1.3). Fichtelius, Johansson and Nordin (1980) isolated intonational features by filtering speech to eliminate segmental information. The result is a signal in which words are no longer recognisable but rhythmic and pitch features of the original speech signal are unaffected. They noted that "the acoustic variable showing the greatest covariance with the respondents' judgement of sex as well as the speakers' actual sex is the number of large frequency variations per time unit" (1980; [223]).

According to the evidence that appears in the literature up to this point, adult male speech can generally be distinguished from adult female speech in that the male pitch range is narrower than the female and shows slower and less frequent pitch shifts. Female speech is probably also associated with more and greater shifts in amplitude. Overall, females can generally be said to be more dynamic in their use of pitch.

In one of the few studies to examine other factors, Robb and Simmons (1990) investigated the vocal fold contact behaviour of 26 children between the ages of 4;4 and 6;6 years. They used an electroglottograph (EGG) to gain information regarding the children's vocal fold opening and closing movements and fundamental frequency. Mean contact quotients[7] (Q_C) for productions of 3 vowels by males and females are shown in table 1.3.

Table 1.3 Mean values of Q_C of 3 sustained vowels by prepubertal males and females (Data from Robb and Simmons (1990)).

	Vowels		
	/i/	/u/	/a/
males	0.56	0.60	0.62
females	0.54	0.57	0.54

Clearly, all of the values of Q_C are greater than 0.50 which indicates that there was more glottal opening than closing behaviour per cycle and, surprisingly, the boys' values were greater than the girls' values although only the values of /a/ were significantly different. Investigations using adults have suggested that females show less vocal fold contact per glottal cycle (i.e. a greater open quotient or, in terms of the Q_C measure, a lower contact quotient) than males. Robb and Simmons' results show that the opposite situation holds amongst their sample of prepubertal children.

[7] Somewhat confusingly, Robb and Simmons (1990) determine Q_C by dividing total closing time by the period

The authors offer the larger intragroup variability of the boys across the vowels compared to the girls as a possible explanation for this result. They point out that "the variability in Q_C among the boys could have been related to vowel specific variations in loudness." (Robb and Simmons, 1990: 1321). Any differences in the phonatory level of the children's utterances might affect the amount of vocal fold contact - Orlikoff (1990) found that Q_C increased with 10 dB increases in vocal SPL. This finding casts doubts on the validity of the interpretation of the results as a true gender difference.

Measures of breathiness of prepubertal children's voices are not reported in the literature.

1.3 EXPERIMENTAL EVIDENCE ON IDENTIFICATION OF CHILDREN'S GENDER FROM VOICE

That gender is identifiable from speech is a forgone conclusion as far as adult voices are concerned. Schwartz and Rine (1968) showed that listeners achieved a correct recognition rate of 100% based only on samples of isolated whispered utterances of the vowel /a/. The majority of studies into adult vocal sex-differences have concentrated instead on determining acoustic parameters which might act as the crucial bearers of gender identity (Lass, Hughes, Bowyer, Waters and Bourne, 1976; Bennett and Montero-Diaz, 1982; Günzburger, 1984).

In contrast, comparatively little work has been done on the identification of prepubertal children from voice. Adult studies have concentrated almost entirely on the relative contributions of the fundamental and formant frequencies to the gender recognition process given that these features are perceptually the most useful cues available to the listener. To recap from section 1.1.2, the fundamental frequency difference between adult males and females is largely due to enlargement of the larynx and the lengthening and thickening of the vocal folds of the male during puberty, however, there is evidence of the existence of cues to the sex of the speaker in the voices of prepubertal children, suggesting that not all speaker sex-identification cues develop at puberty and therefore that some cues may be patterns of acquired behaviour.

Weinberg and Bennett (1971) used spontaneous speech samples from 5-6 year old American children. Listeners correctly identified the gender of these children from the recordings on 74% of all judgements. 20 of the 29 boys were consistently identified correctly as boys and only 2 were misperceived as girls and 25 of the 37 girls were identified correctly as girls with only 3 being misidentified as boys. In order to assess the involvement of fundamental frequency in the ability to discriminate gender from voice the experiment also included an analysis of Fo. Results showed no difference in mean Fo between the boys and girls and Weinberg and Bennett speculate that differences in formant frequencies might be at least partly responsible for the perceptual distinction.

Sachs *et al.*(1973) used sentences recorded from a total of 26 boys and girls between the ages of 4 and 12 years as their experimental stimuli. Judges who listened to these sentences managed an average correct response rate of 81%. 12 of the 14 boys were correctly identified as were 9 of the 12 girls - (2 girls were identified wrongly as boys). The acoustic analyses (which were performed on isolated vowels) showed the boys to have a significantly higher Fo and lower formant frequencies than girls. On the basis of this result, Sachs *et al.* claimed that Fo was probably not used greatly in the perception of gender of prepubertal children. The acoustic cue which appears likely to be of the most use to listeners in identifying the sex of young children is the difference in formant frequencies. In addition to the analysis described above, Sachs *et al.* grouped together the best and worst identified children of each sex. As expected, the most boy-like voices (i.e. the best-identified boys and the worst-identified girls) had lower formants than the least boy-like voices. Sachs *et al.* consider various theories to account for this acoustic difference, including the possibility that culturally acquired norms of male and female speech styles influenced the children's articulations. In other words, they argue that children can deliberately manipulate their pronunciations so as to vary their formant frequencies in a way that conforms to male \ female archetypes.

Alternatively, we have already considered the possibility that boys and girls have anatomical size differences in their vocal tracts prior to puberty which might result in lower formants in boys (see section 1.2.3), although Sachs *et al.* argue against this.

In order to consider the influence of factors other than formants, Sachs (1975) performed two experiments using speech samples recorded by Sachs *et al.* (1973). Firstly, she analysed judges' responses to the productions of three isolated vowels spoken by the prepubertal children (/i/, /u/ and /a/). The mean number of correct responses was better than chance level with 66% of all judgements being correct, however the identification rate from vowels was less than the identification rate from sentences (66% vs. 81%) and Sachs hypothesised that isolated vowels carry only some of the information which allows listeners to correctly judge gender from voice.

Secondly, as the identification rate was lower when listeners heard only vowels as opposed to complete sentences, Sachs devised a method to expose listeners to more phonetic material than is found in vowels but less than a complete sentence. She presented listeners with recordings of the original test sentence of Sachs *et al.* (1973) but played backwards. The rationale for this is that backward speech preserves some acoustic features of the original (e.g. formant frequencies) but masks any information regarding intonation or speech rate. Analysis of the correct response rate indicates that 59% of judgements were correct, this level is not significantly above chance.

Overall, listeners performed most accurately when judging gender from normal sentences and less accurately when making judgements from isolated vowels or backwards sentences. The accuracy of responses to backward sentences is not significantly different from the accuracy of responses to vowels. Sachs (1975) suggests, therefore, "that there may be considerable information in normal sentences that provides a clue to the sex of the speaker, beyond the phonetic aspects of the voice." [160]. Sachs concludes that there are probably a number of cues which together account for the listeners' accuracy including sentential cues and voice quality cues.

Bennett and Weinberg (1979a,b) carried out a comprehensive study of a number of types of utterance using children between the ages of 6;1 and 7;10. Perceptual judgements of sex were obtained in response to recordings of normally phonated and whispered vowels, normally spoken sentences and sentences spoken in a monotonous fashion. These particular types of utterance were chosen to permit the testing of two hypotheses : firstly, that sexual differences in the vocal tract resonance characteristics of preadolescent children provide major cues about sexual identity when judgements are based on isolated vowels; secondly, that cues derived from the intonation patterns of children provide perceptually relevant information about sexual identity.

The results from the perceptual judgement experiment are summarised in table 1.4 overleaf. Statistical analysis of this data revealed significant main effects for speaker sex and utterance type and a significant sex × utterance interaction. Further analysis of the interaction indicated that the average rates of correct identification of gender from phonated and whispered vowels did not significantly differ. This finding held true for both male and female speakers and suggests that sexual differences in vowel formant frequencies provides major perceptual cues about the sex of these prepubertal children. An interesting effect was obtained when the authors analysed the sentential data for gender judgement rates. As is clear from table 1.4, listeners performed with significantly different levels of accuracy for the two sentence types (normal and monotonous) for both the girls and the boys. For the girls, listeners' accuracy in gender judgement was 6% lower when listening to monotonous sentences compared with normal sentences. For the boys however, listeners' accuracy using sentences spoken on a monotone was 9% *higher* than for the normal sentences. The authors summarise these findings thus :

> "It is apparent that monotonicity had a deleterious effect on the perception of femaleness and an enhancing effect on the perception of maleness....This observation suggests that the perception of male vocal quality in the voices of some children may have been influenced by the presence of decreased variations in fundamental frequency....On the basis of this data, it is likely that vocal tract resonance characteristics [....] provided primary information about the sexual identity of a majority of children."
>
> Bennett and Weinberg (1979a; [183])

Table 1.4 Percentage of correct listener identifications as a function of utterance type and child sex. After Bennett and Weinberg, (1979,a : 183)

Utterance type	Mean correct listener identifications	
	Boys	Girls
Whispered Vowels	66.78%	65.02%
Phonated Vowels	67.97%	62.79%
Sentence (monotone)	80.81%	62.99%
Sentence (normal)	71.36%	68.85%
Overall	71.50%	64.90%

Although monotonicity appears to significantly affect listeners' perceptual judgements, intonation patterns did not provide critical information to the listeners in Bennett and Weinberg's experiment. 87% of children received the same perceptual judgement for both their normal and monotonous sentences. The evidence from these experiments that vocal tract resonances provide the critically relevant information about children's sexual identity may be summarised as follows:

- the rates of correct identification recorded for children's whispered vowels was generally very high
- removing suprasegmental cues in the monotone condition did not alter the recognition of most children's gender (because intonation was not the critical acoustic cue)
- F1 and F2 correlated significantly with listener judgements of children's gender in each of the four utterance types
- F1 and F2 together accounted for the largest proportion of the variance in listener perception of gender

Bennett and Weinberg, (1979a,b)

1.4 SUMMARY

The fact that listeners can quite easily distinguish the voices of adult males from adult females on the basis of voice alone reveals that the acoustic cues which underlie this ability must be perceptually prominent. Traditionally, there are considered to be two parameters which contribute to the perception of sexual identity in the voice : fundamental frequency and formant frequencies. For adults, it is generally believed that the fundamental frequency is the more important component which contributes to maleness or femaleness in the voice. Due to a sexual dimorphism which occurs at puberty, adult males have a lower F_0 than females and, in an experiment designed to assess the identification of gender of normal adults from their vowels, Lass, Hughes, Bowyer, Waters and Bourne (1976) found a correct recognition rate of 95-100%. When the same adults produced whispered vowels however, the rate of correct gender identification dropped significantly. In this condition, the listeners were presumably basing their judgements on other parameters including vocal tract resonances.

Formant frequencies are directly dependent on the size and shape of the vocal tract and its cavities. As adult females have smaller vocal tracts than men they will have correspondingly lower formant frequencies, thus "an acoustic cue to the sex of the speaker purely based on vocal tract resonances which result in characteristic vowel formants should be contained in the voice regardless of fundamental frequency" (Günzburger, 1984; [17]). Coleman (1971, 1976) performed a number of experiments in order to assess the relative importance of the fundamental and formant frequencies to the process of gender recognition. He found a very high correlation between the Fo of a voice and the degree to which it was judged to be male or female but a much lower correlation for vocal tract resonances.

Three quotations which relate to similar experiments carried out by three different investigators highlight the differences in theoretical opinion held by different individuals. The first two experiments describe the situation in which a voice is resynthesized with the characteristic Fo of one sex and formant frequencies more suited to the opposite sex. The third experiment concludes that neither Fo nor formants are sufficient vocal cues to gender.

"When a laryngeal vibrator producing a sex-specific Fo is substituted for the normal glottal tone...the male Fo retains its perceptual prominence [over the female vocal tract resonance characteristics]."

Coleman (1976; [179])

"...formant structure is a relatively more important cue in speaker identification than fundamental frequency. For example, vowels produced with male formants, but female Fo were assigned to male speakers in 80.8% of instances, whereas...vowels produced with female formants and male Fo were assigned to males only in 18.6% of instances."

Lehiste and Meltzer (1973; 88)

"...straightforward acoustical parameters like Fo and formant values cannot have functioned as a key to successful sex identification."

Günzburger, Bresser and Ter Keurs (1987; 52)

Chapter 2 Method

2.0 INTRODUCTION

The method of the study fell into two conceptually distinct (if not chronologically distinct) sections. Firstly, the perception experiment, during which data from the child subjects was recorded, prepared and presented to the adult listeners. Gender judgement data was elicited from the adult subjects by means of perception tests and subsequent statistical analyses were performed. Secondly, the acoustic investigation, during which the children's speech recordings were acoustically and statistically analysed

2.1 PERCEPTION TEST

2.1.1 Child Subjects

One hundred child subjects (50 boys and 50 girls) were recruited from two local Edinburgh primary schools. Full ethical approval for the project was granted by the Lothian Regional Council Department of Education and information letters[8] were distributed to parents / children who were then invited to volunteer. Certain children proved to be unsuitable for participation in the study and they were screened at the recruitment stage. These children included those:

- with parental-reported hearing deficits;
- with gross language production problems (e.g. cleft palate);
- who were attending a speech therapist;

One further condition of participation was that the children belonged to families which were local to the Edinburgh region. English was the only language spoken in

[8] The information letter and consent form is included as Appendix 2A

their homes and the children spoke with the local Scottish accent of English. These conditions were ascertained by the researcher.

From the volunteer subjects, those children between the ages of 4.0 and 5.11 (but particularly those between 4.6 and 5.6) who also reported normal hearing were selected to participate.

The methodological techniques and broad purpose of the study were explained to parents and children, but they were not informed of the study's relation to gender as it was felt that this might insert bias into the children's responses.

2.1.2 Data gathering (recordings)

2.1.2.1 Material

The following speech samples were selected as appropriate for the subsequent acoustic and perceptual analyses :-

i. Isolated vowels : The children sustained the vowels /i/, /u/, /a/ and /o/ at a comfortable pitch and level. The vowel /u/ was later excluded from the study on grounds of inconsistency of production[9].
   These particular vowels were selected because it was felt that they represented enough of a contrast between high, low, front and back articulations to highlight any spectral gender difference which may exist.

ii. Isolated sentences : The two sentences chosen were: *Rover is a big, brown dog* and *Rover has got a bone in his mouth.* In order to elicit these sentences from the children without influencing their pronunciation or intonation it was decided

---

[9] The realisations of the four vowel phonemes by the children did not necessarily correspond to the realisations of the same phonemes by the researcher and so there was the possibility of the children varying their responses between a precise auditory copy of the researcher's vowels and their own idiosyncratic pronunciations. This effect was most noticeable in the vowel /u/. Thus, where the researcher pronounced /u/ as [u], some children also pronounced it as [u] (therefore copying the sound as closely as possible), however some children pronounced it as [ʉ] (which was consistent with his/her accent). As the pronunciation of the vowel /u/ varied so greatly across the children, it was excluded from any further part in the study.

not to have them repeat directly the production of the researcher; instead the children were introduced to two colour pictures each of which was identified with one of the test sentences. Once the child had correctly learned which picture accompanied which sentence, the pictures were presented to the child and he or she was prompted to utter the appropriate sentence, thus avoiding verbal cueing.

iii. Spontaneous speech : A period of quasi-spontaneous speech was elicited from each child by means of the 'Bus Story' (Renfrew, 1969). This procedure involves the researcher narrating a story which is illustrated by colour pictures viewed by both child and researcher. The child is then asked to re-tell the story from memory using the pictures. The resulting speech sample has the advantage of being of a very similar topic across all children but leaves the final choice of wording open to each individual child.

2.1.2.2 Data Collection Procedure

All of the recording sessions were carried out within the confines of the two schools. Prior to the beginning of the recording sessions, the researcher spent several hours interacting with the children in various different classroom activities to allow the children to become accustomed to his presence. The child subjects were taken to a quiet room individually to perform their recordings. After a period of time during which the child was put at his or her ease, the speech sample elicitation procedure was begun.

The children were always encouraged to make a game of eliciting speech. In the case of isolated vowels, for example, the researcher used a phrase such as "I'm going to make a sound now. Can you say it back to me?". The vowels were then produced by the child repeating the researcher's productions. In this way, samples of each of the four stimulus vowels were elicited. If the child did not make the same phonetic vowel quality as the researcher at first, he or she was prompted to listen again as the researcher produced the vowel and then to try a second time. If this still resulted in a different quality, the researcher moved on to the next vowel or to the next task.

The two experimental sentences were then elicited using the visual cueing method outlined above.

In order to elicit spontaneous speech from the child, the 'Bus Story' (Renfrew, 1969) was used. The 'Bus Story' is a package developed and used by speech therapists to assess semantic production skills in children, however its use as a method to elicit a sample of spontaneous speech is also widespread. The rationale of the 'Bus Story' is that the researcher / therapist reads a short story to the child concerning the adventures of a naughty bus whilst illustrating the story with a specially provided picture book. After ensuring the child's attention during this phase, the researcher then asks the child to re-tell the story, using the picture book for guidance. In this study, the children's retelling of the story formed a reasonably sized sample of quasi-spontaneous speech. In those cases where the child was not immediately forthcoming in the story-telling process, the researcher prompted the child with a question such as "What's going on in this picture?" or "What did the bus do next?". This method of speech elicitation has the advantage of controlling to a large extent the topic of discourse whilst still giving scope for individual spontaneity.

During the data gathering, the vowels were elicited first, followed by the sentences and, finally, the spontaneous speech. The recording session lasted for approximately 15 minutes per child. As data could only be collected in the mornings due to the constraints of school time-tabling, the recording sessions were conducted over a period of several weeks.

2.1.2.3 Recording Equipment

Speech data was recorded in a number of relatively quiet (but not sound-proofed) rooms in the participating primary schools (e.g. Medical room, Head's office, Library etc.). Recording equipment consisted of a Realistic® PZM Microphone and an AIWA HD-S1 Digital Audio Tape recorder fitted with a HDA-1 A/D converter. The data was recorded directly onto Maxell DM90 DAT tapes at a sampling rate of 48kHz with 16-bit Digital to Analogue (quantization) rate.

2.1.3 Editing

The master tapes, on which were recorded all of the child subjects' data, were edited in such a way as to group all of the childrens' utterances of each type of speech sample together, that is, vowel repetitions for all of the children were grouped together on one tape, sentence repetitions for all of the children were grouped together on a second tape and spontaneous speech for all of the children were grouped together on a third tape. The order of recording of the child subjects was randomised for each sample type, however within the vowel and sentence tapes all of each individual child's utterances remained together. Thus there were three edited tapes: tape 1 consisted of vowels; tape 2 consisted of sentences and tape 3 consisted of spontaneous speech.

As discussed above, the vowel /u/ was not included in the final edited vowel tape because it suffered from a variable realisation across the pool of child subjects.

Of the 100 productions of /u/, 32 were transcribed by the author as [o], 4 were indeterminate and 64 were recognisable as [u]. As the vowel /o/ was also included in this study and as it was realised consistently as [o], it was felt that it would be confusing to include a phoneme /u/ which was incorrectly realised as [o] in a third of its productions. The vowel /u/ is also prone to fronting in accents of Scottish English and the perceptual difference between the author's /u/ realised as [u] and some of the children's /u/ realised as [ʉ] may have been a further source of confusion.

From the original pool of 100 children a final sample of 89 children was selected (46 males and 43 females). The 11 children who were excluded from the experiment at this stage were screened out either because their speech was so quiet as to make their recording inaudible or because of abnormally poor speech production which rendered their speech incomprehensible. Each tape contained all 89 child subjects in a different random order with each individual child's data remaining together. All editing was performed using SONY DTC-690 digital audio cassette recorders and Maxell DM-90 DAT tapes.

2.1.4 Presentation of Stimuli / gender Judgements

2.1.4.1 Adult Subjects

The adult subjects were members of the academic staff and students of Queen Margaret College, Edinburgh who volunteered to assist with the experiment. These subjects were each paid £10 for their participation in the study. Members of the Department of Speech and Language Sciences were excluded from the panel of subjects as it was felt that their specialist training in speech perception might prejudice the outcome of the listening experiment.

The panel of listeners consisted of 8 male and 8 female subjects with an age range between 18 and 33 years. All adult subjects reported themselves to be native speakers of English and to be free of hearing disorders. The adult subjects were given reference numbers of M1-8 (males) and F1-8 (females).

2.1.4.2 Material

The data provided by the adult subjects was in the form of gender judgements of each of the three types of speech samples spoken by each of the child subjects.

All of the adult subjects judged all of the children's speech samples. The adult judges recorded their responses on an answer sheet by circling 'M' or 'F' (for boy and girl respectively) in the appropriate box. Each judge received one answer sheet which was divided into three sections (one section for each type of speech sample - vowels, sentences, spontaneous speech).

All of the vowels for one child (/i/, /a/, and /o/) represented one vowel stimulus, similarly both of the sentences (*Rover is a big, brown dog* and *Rover has got a bone in his mouth*) represented one sentence stimulus and the passage of spontaneous speech represented one spontaneous speech stimulus; that is, the judges made one response to each child's set of vowels, one response to each child's set of sentences and one response to each child's spontaneous speech. Therefore there were 3 gender responses to each child. As there were 89 children, this realised a total of 267 gender judgements per judge and a grand total of 4272 gender judgements in the whole experiment.

2.1.4.3 Data Collection Procedure

Adult subjects, like the child subjects, were asked to volunteer for the study and having successfully recruited 16 subjects they were gathered together at a mutually convenient time to listen to the child subjects' data and to make their responses.

The purpose of the experiment was explained to the judges before beginning the task and they were each given written and oral instructions on how to record their responses. The adult subjects were seated in a relatively quiet room and listened to each tape through headphones which connected directly to a SONY DTC-690 digital audio cassette player. Judges marked their estimate of the gender of each child on the response sheet[10].

A short pause was included between tapes to allow the judges to rest. The gender of each judge was recorded along with their responses to permit analysis of the effect of the sex of the listener on gender identification.

2.1.5 Statistics

The judges 'male' and 'female' responses to each child were collated and entered into a spreadsheet (Microsoft Excel™) where totals and means were calculated. In addition to the total number of male (boy) and female (girl) responses made by male and female judges, the total <u>correct</u> responses was calculated. These figures were converted into percentages giving a clear numerical indication of the relative rates of accuracy of both male and female judges and also of how well the boys and girls were identified with respect to each other.

[10] See appendix 2

Single t-tests (paired or independent) could be used to assess the likelihood that the differences between how well boys and girls were identified or how well male and female judges performed for each sample type. However, in order to achieve full coverage of all possible comparisons, at least six different tests would have to be carried out. If a t-test is performed with a significance level of 0.05 then there is a 5% risk of a Type I error (false rejection of a null hypothesis) occurring. However if multiple comparisons are performed, the chance of Type I errors increases rapidly with an increasing number of means to compare. With 6 means to compare the chance of a Type I error is 0.37 (Samuels, 1989).

Given the vulnerability of the mutiple comparisons design to Type I error, the data in the present experiment was analysed using the analysis of variance technique.

Two 3-way analyses of variance with repeated measures were used to assess differences between the correct recognition rates. In the first ANOVA the between subjects variable (grouping) was sex of judge which had 2 levels (male, female) and the within subjects variables (repeated measures) were sex of child (also with 2 levels - male, female) and material type (which had 3 levels - vowel, sentences, passage). In the second ANOVA, the between subjects grouping variable was sex of child whilst the sex of judges and material type were within subjects repeated measures.

Following the discovery of a significant F-ratio for the material type variable, it was necessary to determine which of the three means differed from each other in order to complete the analysis. A post-hoc analysis was carried out using the Scheffé test.

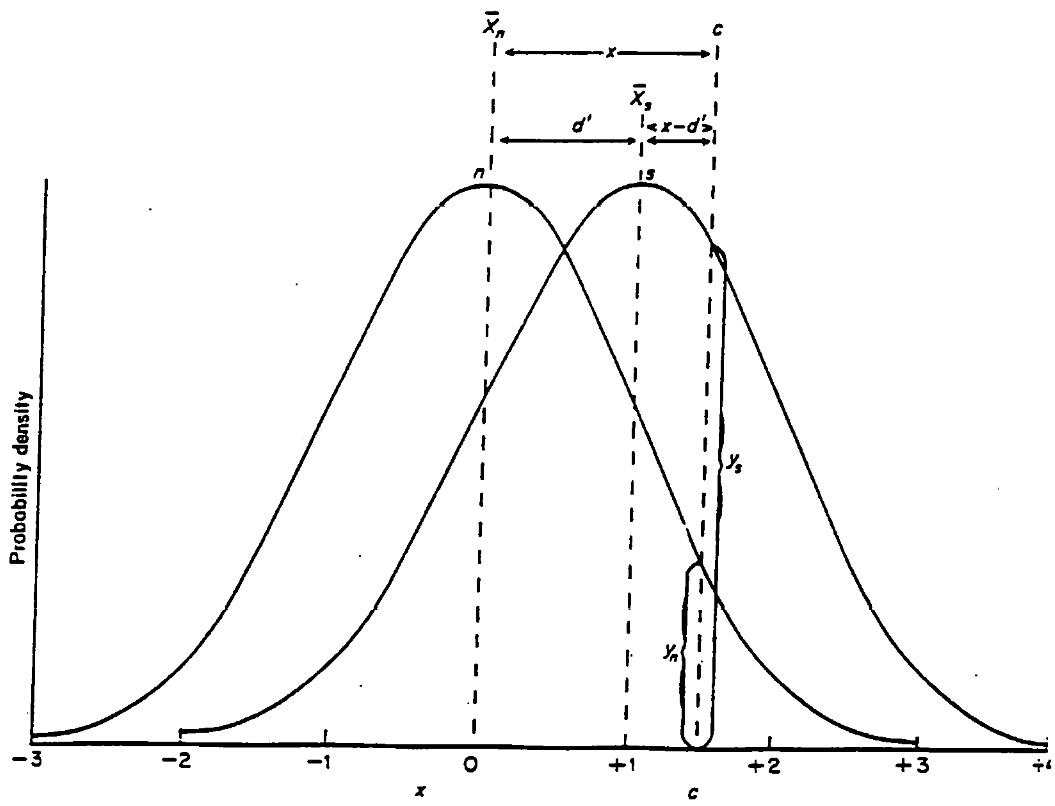
2.1.5.1 Bias and Discrimination

It is possible to assess the prejudice of each individual judge to respond male or female across the entire set of stimuli. The ideal judge would of course be no more inclined to respond male than female when averaged across an equal number of male or female stimuli and such a judge's responses could therefore be taken to represent only the influence of that individual's gender recognition ability. Any discernible bias which a judge might bring to the task would be reflected in an unusual distribution of the correct responses ('hits') and incorrect responses ('false alarms'). Signal detection theory provides the means to quantify bias measurements and allows the evaluation of the judges' ability to discriminate gender.

Discrimination

Consider figure 2.1 which shows a signal (s) and a noise (n) distribution both of which are distributed normally. The judge's ability to discriminate signal from noise is called d' (d-prime) and is represented here as the distance of the signal mean from the noise mean measured in multiples of the noise standard deviation. For convenience, the mean of n is taken to be 0 and its standard deviation is taken to be 1.

Figure 2.1 Gaussian distributions of signal and noise and a single criterion, c. Original figure is from McNicol (1972).



In other words, d' is a standardized or z-score. In order to calculate d', the hit and false alarm rates for each judge must be determined. Normally this would involve finding the proportion of the area under a standard normal curve which lies to the right of a point on the x-axis x S.D. units from the distribution mean. In our case, however, we can compute the hit and false alarm rates directly from the raw data according to the table below[11]

	RESPONSES		TOTAL STIMULI
STIMULI	S= "boy"	N= "girl"	
s (boys)	**HIT** P(S|s) = Σ'boy' reponses / Σboy samples	**MISS** P(N|s) = Σ'girl' reponses / Σboy samples	Σ boy samples
n (girls)	**FALSE ALARM** P(S|n) = Σ'boy' reponses / Σgirl samples	**CORRECT REJECTION** P(N|n) = Σ'girl' reponses / Σgirl samples	Σ girl samples

As d' represents the distance between the mean, $\overline{X_s}$, of a signal distribution and the mean, $\overline{X_n}$, of a noise distribution we can define d' as

$$d' = \overline{X_s} - \overline{X_n}$$

As we stated above, d' is measured in standard deviation units of the noise distribution and so the formula should more correctly be defined as

$$d' = \frac{\overline{X_s} - \overline{X_n}}{\sigma_n}$$

where σ_n is the standard deviation of the noise distribution. The reader will notice the similarity between this formula and the formula used to convert a raw score into a standard score (standard normal deviate). This is the rationale behind thinking of d' as a z-score. The working formula therefore for finding d' from a pair of P(S|n) and P(S|s) values is

[11] In this case I have arbitrarily assigned the response 'boy' as signal and the response 'girl' as noise. It should be noted however that if the judges ability to discriminate females was the focus the boy / girl assignments should be switched.

$$d' = z(S|n) - z(S|s)$$

where z(S|n) is the z-score corresponding to P(S|n) and z(S|s) is the z-score corresponding to P(S|s).

Simple d' scores were calculated for each judge in each speech sample type. The values of d' for each sample were compared in order to determine the statistical difference between the performance (accuracy) of judges when listening to different speech types. In this analysis sign tests were used to assess the differences between d' scores for all judges (n=16).

<u>Bias</u>

Using the same theory of signal detection, we can learn something about the judges' response bias. The statistic which we use to measure this bias is β (beta). The construction of the formula for β from first principles is somewhat complex and the interested reader is directed to McNicol (1972) for a fuller account than is presented here. Looking again at fig 2.1 we have seen that the distance between the means of the two distributions shown is equal to d'. Note that a criterion, C, is situated at a distance of $+x$ from the noise distribution mean. The value of β associated with this criterion is found by dividing y_s, the height of the signal distribution at C by y_n, the height of the noise distribution at C. In practice, y values for a range of P values can be read off statistical tables and all that is necessary is to substitute them into the formula to calculate β.

$$\beta = \frac{y(P(S|s))}{y(P(S|n))}$$

At the point where signal and noise distributions intersect, β =1. At this point therefore the judge is no more biased to S ('boy') than N ('girl') responses. To the left of this point on the graph the noise distribution lies above the signal distribution and β is less than 1. In this area the subject is biased towards 'signal' responses. To the right of the point of intersection β>1 and the subject is biased towards 'noise' responses. β can thus be seen to be a measure of response bias of any given criterion point.

In terms of this experiment β represents the predisposition of a particular judge to respond 'male' or 'female'.

β	**Criterion**	**Bias**
$\beta<1$	lax or risky criterion	bias towards 's' (male)
$\beta=1$	unbiased	no bias
$\beta>1$	strict or cautious criterion	bias towards 'n' (female)

The reader should exercise caution when interpreting β values as criteria which represent biases towards 's' (male) are restricted to the narrow range $0 < \beta < 1$ while criteria which represent biases towards 'n' (female) can take any value $\beta>1$. It is easy to be deceived by the amount of bias shown in a β value. For example, $\beta=2$ represents the same degree of bias to 'n' as $\beta=0.5$ represents to 's', while $\beta=100$ represents the same degree of bias to 'n' as $\beta=0.01$ represents to 's'. To normalize the distances between response biases and to allow for an easier graphical layout, bias scores are often presented in terms of log β rather than β (McNicol, 1972).

Where there is bias towards 's' responses log β will be positive, when there is bias towards 'n' responses log β will be negative and when there is no bias log β will be zero. Furthermore, the strength of the bias will be in proportion to the distance of the log β value from zero. Therefore, in this experiment, if $\log \beta = 0.2$ we can say that there is a bias on the part of that judge to respond 'female' whereas a value of $\log \beta = -0.35$ would represent a relatively greater degree of bias on the part of the judge to respond 'male'.

2.2 ACOUSTIC ANALYSIS

Due to the nature of the different speech samples, different analysis techniques were used to extract information from each. The sentences and vowels represented the sample types which yielded the best and worst identification rates respectively by the judges and so these were selected for acoustic analysis. As there was no significant difference between the recognition rates yielded by sentences and spontaneous passage, and also because of the difficulty involved in extracting stable acoustic measures from fast, coarticulated passages of child speech, there was no acoustic analysis performed on the spontaneous speech samples.

2.2.1 Vowel Sample

The children's vowels were digitized at a sampling rate of 22.05kHz (16 bits / sample) and stored on hard disk.

2.2.1.1 Extraction of Vowel formant information

Overview of Multiple Centroid Analysis

In acoustic research of the past twenty years if an experiment required the extraction of formant information from speech samples it would very likely be achieved by either of two methods: manual estimation of spectrographs or Linear Predictive Analysis (LP).

Both of these methods however have associated drawbacks. The process of estimating formant frequencies from spectrographs is laborious and intensive and is prone to measurement error. Furthermore, formant bandwidths and amplitudes are practically unrecoverable by means of simple visual inspection. LP, on the other hand, can be performed automatically by computer and yields full information on frequency, amplitude and bandwidth of a given formant. LP analysis, however, copes badly with high-frequency speech (i.e. women and children). It works by solving for the roots of the linear predictor polynomial and using a tracking algorithm to assign the correct root to the correct formant. The major problem for high-pitched voices is that the formants are linked to spectral peaks and the higher the fundamental frequency, the more widely spaced are the harmonics. If a vocal tract resonance happens to fall between two widely spaced harmonics, there will be considerable error in the location of the formant peak.

Due to these limitations in the available techniques for formant estimation, a relatively new method was adopted - the **constrained Multiple Centroid Analysis** (MCA).

In an experimental comparison using child subjects MCA proved to be faster and more accurate in formant extraction than LP and it showed less bias towards the nearest harmonic (Wrench, Watson, Soutar, Robertson and Laver, 1994). If the researcher is interested in extracting 2 centroids, for example, then the basic centroid analysis involves sectioning the frequency distribution of a sample into 2 nonoverlapping partitions and assigning one unique centroid in each. The centroid is located at the frequency that yields the minimum squared error value. Squared error is given as:

$$\sum_{n=k_1}^{k_2} (n-k)^2 P(n)$$

The minimum squared error values are summed and stored for each possible partitioning of the sample. The final centroid estimate is the centroids which correspond to the partitioning with the lowest minimum error score (Wrench, 1995).

The output of the analysis can be seen as the overlaying of a normal distribution on the distribution of the power spectrum. If the partition has been correctly assigned then that part of the spectrum within a single partition should contain a single formant. The centroid and its variance will then represent the formant frequency and bandwidth.

The global least squares method used in MCA is computationally expensive and so any procedure for reducing the processing load is advantageous. Because the human vocal tract and articulators have physical limits, there is a restricted range of values in which the formant frequencies can fall. These limits can be coded as constraints on the possible positions of the partition boundaries in the centroid estimation. This technique effectively reduces the combination of possible partitions without losing spectral information and thus represents a saving in computation (Wrench *et al*, 1994).

Procedure

A semi-constrained MCA which searched for up to 6 centroids was implemented on the children's isolated vowel data. The program attempted to locate a maximum of 6 formants in the frequency range from 0 to 6,500 Hz. The lowest three centroids were generally adopted as the frequencies of F1, F2 and F3. Using this procedure, spurious centroids which did not reflect formants tend to cluster around the top end of the frequency range and are therefore easily spotted and separated from the true vocal tract resonances. Formant bandwidths and amplitudes were also extracted from the MCA output.

2.2.1.2 Breathiness

Two measures of breathiness were performed on all of the children's vowels. A very dominant first harmonic (H1) has been widely found to be highly correlated with breathy phonation and so the first measure reflected the difference in amplitude between the first and second harmonics of the voice spectrum. This value is therefore an approximation of the spectral tilt of the vowel, although as the analysis was carried out on the spectrum of the speech output and not on the glottal source spectrum, the influences of the vocal tract transfer function will affect the values. The amount of influence on the spectral tilt depends on the vowel quality; phonetically close vowels will be affected more than open vowels.

A 512 point FFT was carried out at the midpoint of each vowel using the Kay Elemetrics Computerised Speech Lab (CSL). The amplitudes of the first and second harmonic peaks were logged manually and this data was entered into an EXCEL™ spreadsheet. The value of H1-H2 for each individual vowel and the mean and standard deviations of the H1-H2 values for the three different vowels as spoken by boys and girls were calculated.

A number of obvious outliers were apparent upon visual inspection of the H1-H2 data (see fig. 3.14 in section 3.2.1.2 and these were removed from the analysis and the subsequent means and standard deviations were recalculated (fig. 3.15 in section 3.2.1.2).

The reader should be aware of the drawbacks of this type of analysis. Whilst it has the advantage of not requiring the complicated and sometimes questionably reliable technique of inverse filtering, the major problem is that it is sensitive to other factors which can affect the spectral output. The comparison of H1 and H2 levels is only a valid measure when Fo and F1 are reasonably well separated (e.g. in open vowels) - when Fo is high (as is the case in all of the children's voices) and F1 is relatively low (e.g. in close vowels), the amplitudes of H1 and /or H2 may be boosted by the influence of the nearby F1 peak. In this experiment therefore, the H1-H2 results from vowels other than /a/ are affected by filter as well as by source factors and are therefore not valid indicators of breathiness.

The second measure of breathiness involved the comparison of the levels of Fo (H1) and F1.

The Fo amplitude data was the same as in the above measure. The levels of the first formant were automatically extracted from the Multiple Centroid program (see section 2.2.1.1)

Clearly this method takes into account, to a certain extent, the influence of the formant peak on the spectral slope, however there are still difficulties in making cross-vowel comparisons as the relative strength of the first formant peak will depend in part on the frequency location of the formant. Vowels with lower F1 frequencies will tend to have higher F1 amplitudes because of the effect of the 12 dB/ octave spectral tilt. Formant amplitude levels are also partially controlled by the amount of acoustic damping. One characteristic of the breathy voice is that there is a relatively large degree of damping due to the small closed phase duration. In the context of this study, formant damping is a confusing factor as it affects the levels of the output spectrum in a way which is not predictable from the slope of the source spectrum.

2.2.2 Sentence Sample

The recordings of the children's sentences were played into the Kay Elemetrics Computerized Speech Laboratory (CSL), an audio processing package with a number of analysis and display functions, and stored as digitised speech files with a sampling rate of 25KHz. 89 children took part and there were 2 sentences for each child - this therefore yielded a total of 178 speech files. A total of 19 voice parameters were drawn from the speech files using the Kay Elemetrics Multi-Dimensional Voice Program (MDVP). The MDVP works in conjunction with CSL's built-in voice analysis capabilities to perform a multiple parameter extraction on either running speech or steady phonations. The acoustic parameters selected for extraction can be grouped into three sections: fundamental frequency information; amplitude perturbation information and frequency perturbation information. The individual parameters are listed below[12].

- Fundamental Frequency Information — (Average Fo, Average Pitch Period, Highest Fo, Lowest Fo, Standard deviation of Fo, Phonatory Fo Range, Number of Pitch Periods computed)

- Short and Long Term Frequency Perturbation Information — (Absolute Jitter, Jitter Percent, Relative Average Perturbation, Pitch Period Perturbation Quotient, Smoothed PPQ, Fundamental Frequency Variation)

- Short and Long Term Amplitude Perturbation Information — (Shimmer in dB, Simmer Percent, Amplitude Perturbation Quotient, Smoothed APQ, Peak Amplitude Variation)

Frequency perturbation information (jitter) is more often extracted from sustained isolated phonations than from passages of running speech. The rationale for this situation is that many researchers feel that the systematic changes in Fo brought about by voluntary adjustments of the vocal organs for the purposes of stress, intonation and other stylistic effects tend to mask the random perturbations which are of interest. However, the extraction of a jitter measurement from the so-called steady-state portion of a child's vowel is not as straightforward as many would like to make out. Firstly, the sustained vowels of children tend to have less acoustic energy, less duration, a greater degree of nasality and a greater amount of pitch dynamism than the sustained phonations of adults (Kent and Read, 1992). Whilst asking an adult subject to sustain a vowel at a steady pitch, loudness, clarity and for a specific amount of time is not usually a problem, children tend to suit themselves as to their degree of co-operation. The result is that the isolated vowels of children (particularly relatively young children as in this experiment) are often of a highly variable and dubious quality.

The choice of connected speech material therefore as the source of the jitter measurements can be justified on several counts. Using a sample of speech longer than a fraction of a second has the advantage that a longer term average assessment of jitter can be extracted from the children's speech without having to re-perform the analysis over successive windows. This will, of course, have the consequence that the mean jitter values will all be considerably larger than the published normative thresholds although any sex-differences, if present, should be retained and possibly magnified.

It was often necessary to selectively edit the speech files as the MDVP would occasionally misidentify voiceless sections of the waveform[13]. Voiceless fricatives (in particular high energy segments such as /s/) were often mistaken for voiced sections of speech and, on these occasions, it was necessary to fully or partially delete the relevant sections of the waveform from the sample. Furthermore, stop bursts or periods of strong aspiration were occasionally assigned an abnormally high Fo and, because this would adversely affect the validity of the pitch statistics, they would also have to be deleted. The resulting speech files therefore did not contain all of the acoustic information which was available to listeners, however the numerical results extracted by the MDVP should still have accurately represented the values of

---

[12]See appendix 3 for a full description of each voice parameter

[13]The MDVP uses an adaptive time-domain pitch-synchronous method for pitch extraction. See appendix 3 for a detailed description of its operation.

the relevant voice parameters over the whole sample as the sections which were deleted would have given rise to spurious "outliers" in the data.

Once the files were prepared, the above acoustic parameters were automatically extracted from the speech and saved as individual result statistic files. There was one file per sentence stimulus and therefore a total of 178 statistic files. Four files had to be excluded from the final analysis as the results obtained from the MDVP were either clearly spurious or non-existent.

The statistical output was then formatted and prepared for insertion into a spreadsheet. Following this stage, the files were collected into a Microsoft Excel file and preliminary statistics were extracted.

Chapter Three
Results

Section 3.1 describes the results obtained from the perceptual gender-judgement experiment and includes summary statistics. Section 3.2 describes the results of the acoustic analysis and includes averages of the acoustic parameters displayed by actual gender and by perceived gender in addition to other summary statistics.

3.1 PERCEPTION TEST RESULTS

3.1.1 Raw Judgement Data

Table 3.1 shows the proportion of correct gender responses made by judges expressed as a percentage of all responses for male and female children and judges over the three sample types. The differences between these correct gender judgement rates vary between the three speech sample types, the sex of the child and the sex of the judge. Table 3.2 shows the actual number of 'boy' and 'girl' responses to male and female children by male and female listeners and the totals for each group expressed as a percentage of the total number of responses made. It can be seen that the relative number of responses for each sex of listener to each sex of child tally with the 'correct identification' rates in Table 3.1.

Table 3.3 summarises the statistical significances yielded by paired two-samples t-tests carried out between the correct judgement rates of each speech sample type.

Table 3.1 Mean correct recognition rates for male and female children by male and female listeners across vowel, sentence and spontaneous speech conditions.

VOWELS

	Male Judges	Female Judges	AVERAGE
Male children	59%	67%	62.77%
Female children	68%	70%	69.04%
AVERAGE	63.36%	68.45%	65.91%

SENTENCES

	Male Judges	Female Judges	AVERAGE
Male children	71%	75%	73.00%
Female children	80%	79%	79.36%
AVERAGE	75.43%	77.03%	76.23%

SPONTANEOUS SPEECH

	Male Judges	Female Judges	AVERAGE
Male children	67%	74%	70.38%
Female children	77%	73%	74.85%
AVERAGE	71.81%	73.43%	72.62%

Table 3.2 Total number of 'boy' and 'girl' responses made by male and female listeners to male and female children across vowel, sentence and spontaneous speech conditions.

VOWELS

	Male Judges		Female Judges		TOTALS	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
Male children	216	152	246	122	462	274
Female children	110	234	103	241	203	485
TOTALS	326	386	349	363	665	759
Percentage of total responses	45.8%	54.2%	49%	51%	46.7%	53.3%

SENTENCES

	Male Judges		Female Judges		TOTALS	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
Male children	261	107	277	91	538	198
Female children	69	275	73	271	142	546
TOTALS	330	382	350	362	680	744
Percentage of total responses	46.3%	53.7%	49.2%	50.8%	47.8%	52.2%

SPONTANEOUS SPEECH

	Male Judges		Female Judges		TOTALS	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
Male children	245	123	273	95	518	218
Female children	79	265	94	250	173	515
TOTALS	324	388	367	345	691	733
Percentage of total responses	45.5%	54.5%	51.5%	48.5%	48.5%	51.5%

Table 3. 3 Results of paired 2 samples t-tests between data from different speech sample types

	Sentence	Passage
Vowel	$p<0.02$	$p=0.08$ (N.S.)
Sentence		N.S.

The gender judgement performances of the male and female judges, the differences in recognition rates due to the sex of the children and the effects of speech sample type were measured using 3-Way Analyses of Variance with repeated measures. The results of these statistical analyses are summarised in the following sections.

3.1.1.1 Differences in accuracy of listener judgements for girl and boy speakers

Table 3.4 shows the results of the ANOVA in which the gender recognition data is arranged 'by children'. The overall difference in accuracy of gender judgement of boys and girls shows a marginal failure of significance ($p=0.087$).

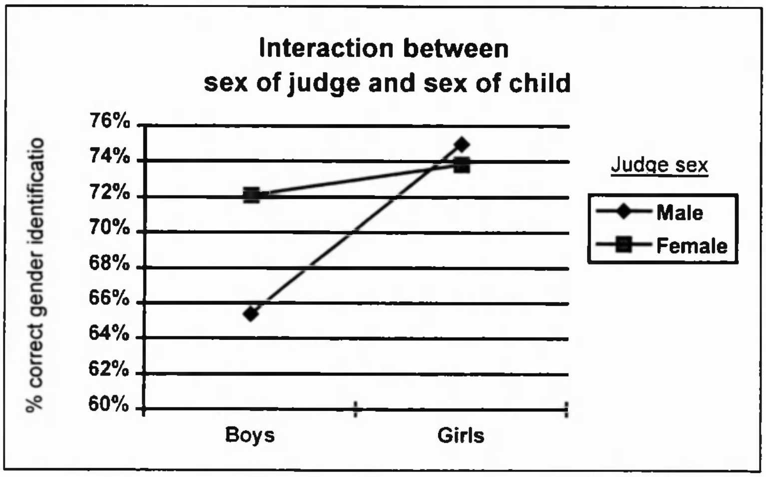
Table 3.4 Results of a 3-way Analysis of Variance with sex of child as the grouping (between subjects) variable and sex of judge and material type as repeated measures.

Source	Sum of Squares	D.F.	Mean Square	F	Tail Prob.
Main Effects					
Sex of child	0.428	1	0.428	3.00	0.087
ERROR	12.429	87	0.143		
Sex of judge	0.102	1	0.102	6.33	0.014
ERROR	1.405	87	0.016		
Material type	0.976	2	0.488	3.68	0.027
ERROR	23.091	174	0.133		
Interactions					
Judge x Child	0.206	1	0.206	12.77	0.001
Material x Child	0.010	2	0.005	0.04	0.965
Judge x Material	0.036	2	0.018	1.00	0.369
Judge x Material x Child	0.028	2	0.014	0.79	0.456
ERROR	3.119	174	0.018		

The main effects of sex of judge and material type were both significant at the 5% level. Sex of child marginally failed to achieve significance. In addition one of the interactions, sex of child x sex of judge, was significant. Figure 3.1 displays the relevant means for the interaction and clearly shows the interaction effect. Whereas the female judges are much better than the male judges at correctly identifying boys, there is little difference between the two groups of judges for the girls (in fact, the male judges are slightly better). Male judges correctly recognised the gender of girls significantly better than boys (p=0.007). Female judges did not recognise either boys or girls significantly better than the other (p>0.05).

A post-hoc Scheffé test was carried out to attempt to determine which of the levels of the materials type variable differed significantly. The results of the test are shown in appendix 4. The Scheffé failed to determine signifcance at p<0.05, although it indicated that the difference between the vowel and sentence samples was the closest to being significant.

Figure 3.1 Chart showing interaction effect of sex of judge vs. sex of child



The number of corect judgements made to individual male and female children can be seen in appendix 5. The reader will note from a brief visual inspection of these figures that the rates of correct identifiction of the sentence and passage samples are generally higher than those of the vowel sample. Notice also that whereas for some individuals a low correct judgement score on one type of speech sample appears to accompany low scores on the other two samples (e.g. boy k20, girls c28, c32), other children score highly on certain samples and poorly on others (e.g. girl c51, boys k10, k29)[14]. There is therefore no simple relationship between correct judgement rates across different sample types. As most of the children were generally well identified in all of the samples, there is a certain degree of natural correlation between the sample scores and it has not been possible to clearly predict accuracy of judgement in any individual sample type given the score from one of the others.

3.1.1.2 Differences in gender judgements made by male and female listeners

The results in table 3.5 give the output of a 3-way ANOVA using sex of judge as the grouping variable and the other variables (sex of child and material type) as repeated measures. The decision to perform two ANOVAs was based on the fact that the variance is different depending on how the ANOVA is designed. The same data is involved but as the ANOVA is centrally concerned with describing the variability within groups of means and between groups of means and comparing the two measures, the outcome of the analysis will vary as a function of the size and variance of each group. The finding that there was no significant difference in the mean performance accuracy of male and female judges when measured across all children can be partly explained by the interaction effect seen in figure 3.1. In other words, although female judges were much better than male judges at identifying boys, the situation was reversed (although with a weaker effect) when girls were being identified. As the ANOVA calculates the *overall* effect and takes into account all of the children when producing an F-ratio for effect of judge sex, the opposite influences of boys and girls cancelled each other out and the result was no significant sex difference in the performance of the judges.

Once again, a Scheffé test was performed on the levels of the material type variable. The Scheffé indicated that sentences were identified at a significantly higher rate than the vowels ($p<0.05$) and the passage was probably also identified at a better rate than the vowels although this effect was not significant (see appendix 4).

---

[14] The prefix ‘c’ or ‘k’ denotes the code for the school which the children attended. ‘c’ was Corstorphine Primary School and ‘k’ was Carrick Knowe Primary School.

Table 3.5 Results of a 3 - Way Analysis of Variance with sex of judge as the grouping (between subjects) variable and sex of child and material type as repeated measures.

Source	Sum of Squares	D.F.	Mean Square	F	Tail Prob.
Main Effects					
Sex of child	0.077	1	0.077	3.16	0.097
ERROR	0.341	14	0.024		
Sex of judge	0.18	1	0.018	1.17	0.297
ERROR	0.220	14	0.016		
Material type	0.176	2	0.088	11.67	0.0001
ERROR	0.211	28	0.008		
Interactions					
Judge x Child	0.037	1	0.037	1.52	0.238
Material x Child	0.002	2	0.001	0.26	0.776
Judge x Material	0.006	2	0.003	0.43	0.655
Judge x Material x Child	0.005	2	0.003	0.76	0.476
ERROR	0.093	28	0.003		

The relative make-up of the pool of subjects by gender is as follows - Males: 51.7%, Females: 48.3%. These percentages can be compared against the percentages of 'male' and 'female' responses in Table 3.1 to show a bias in the distribution of responses (The results of a statistical analysis of bias are presented in section 3.1.2).

3.1.2 Bias and discrimination

3.1.2.1 Raw measurement results

Bias (β)[15] and discrimination (d') measures were calculated for each judge across each sample type. In figures 3.2, 3.3 and 3.4 the values for log β and d' are shown on the same graphs for the vowel, sentence and passage samples respectively. Note that the left *x*-axis shows the log β scale whilst the right *x*-axis shows the d' scale. The

[15] Throughout this section the bias measure will be referred to as log β(male) to indicate that the statistic recognises a response of 'male' as an instance of a signal stimulus and 'female' as a noise stimulus. In fact the corresponding values of log β(female) would be identical with the opposite signs.

reader should also be aware that the scales of these graphs are not consistent across separate graphs. Figures 3.2a and b (which concern the vowel sample) show small positive values of log β for both male and female judges which indicates a small preference to respond 'girl' for most judges. The values of d' are all greater than 0 with one exception and are spread across a range of values. This suggests that different judges show a range of gender-identification abilities, with some being very accurate and others less accurate, but, in general, judges are performing efficiently.

Figure 3.2a Log β(male) and d' for male judges and the vowel sample

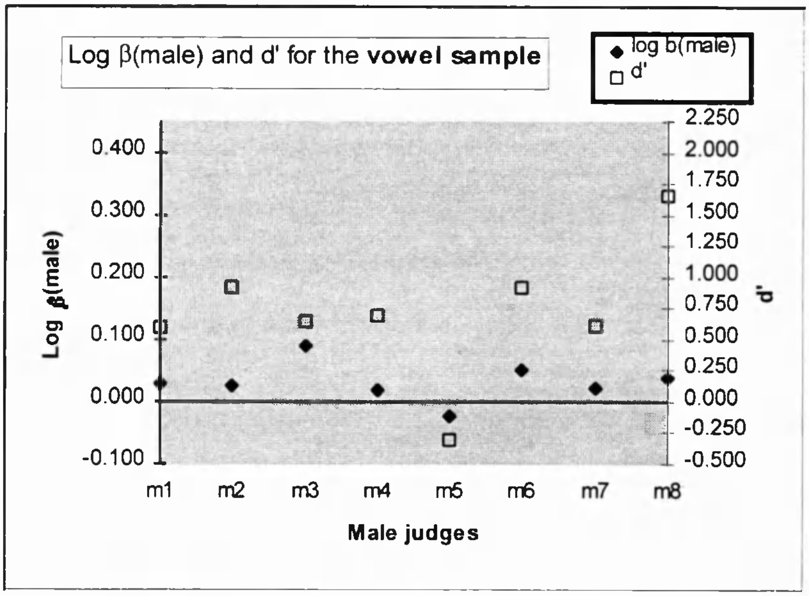
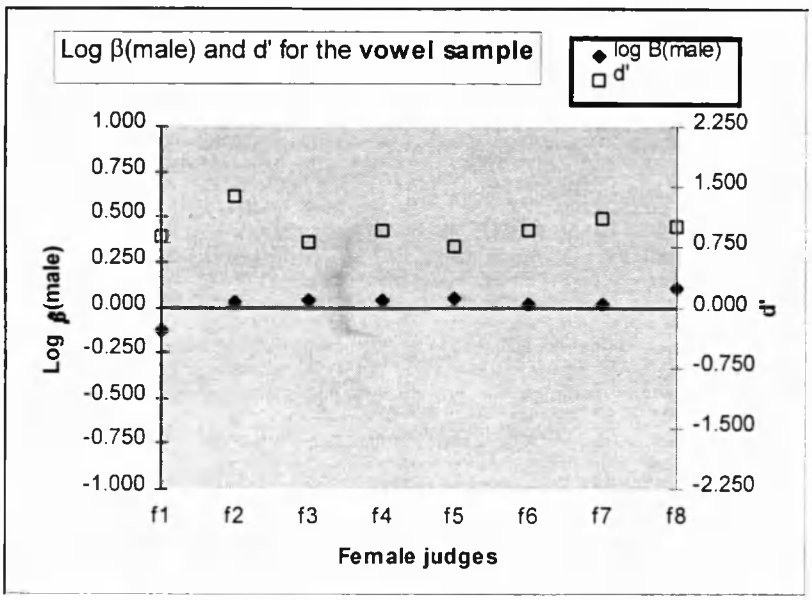


Figure 3.2b Log β(male) and d' for female judges and the vowel sample



In figures 3.3a and b we can see a similar pattern of results as in the previous graphs. These graphs reflect the judges' performances on the sentence sample and we can see that there is an increase in the values of d' indicating that judges are performing more accurately for this sample type. Values of log β are again mostly small and positive with the exception of the score of judge f1. This judge has an extremely large negative value of log β which indicates that she was strongly biased to respond 'boy'. This value is marked as 'outlier' on figure 3.3b.

Figure 3.3a Log β(male) and d' for male judges and the sentence sample

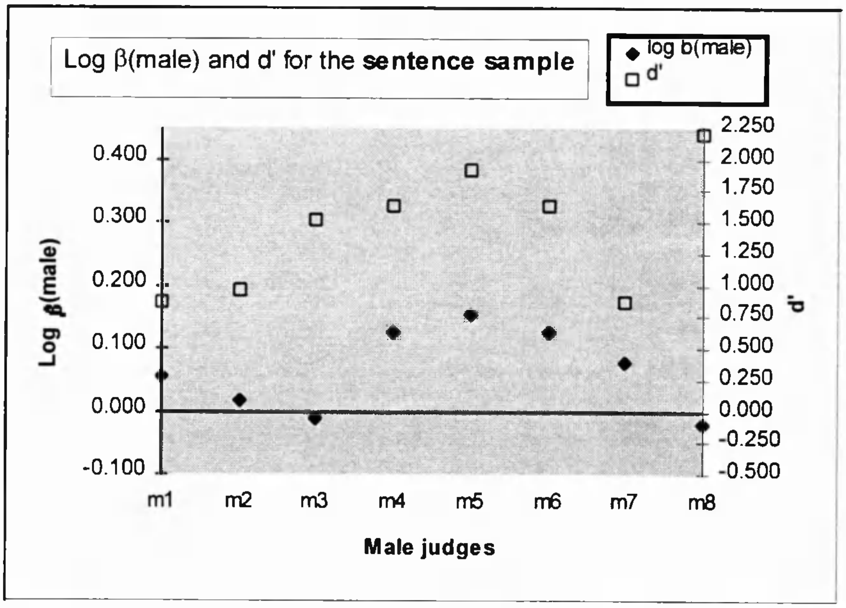
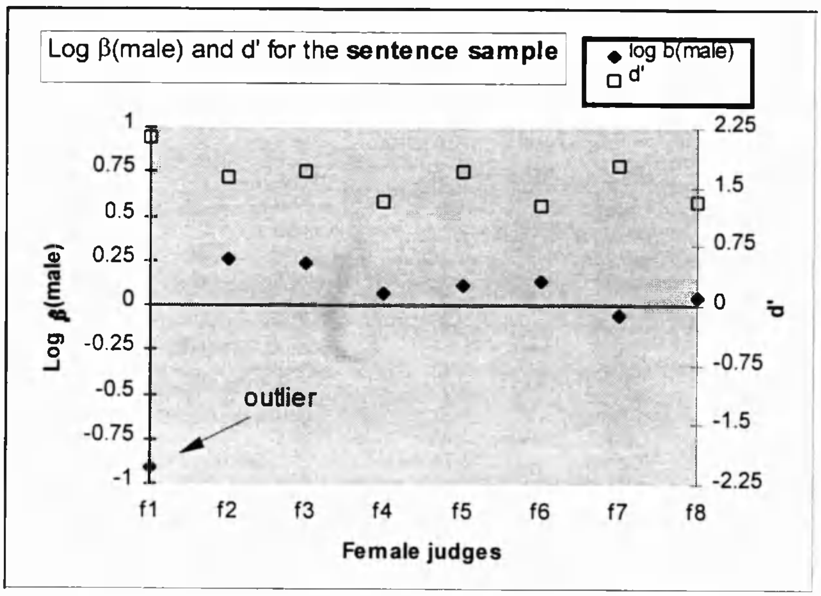


Figure 3.3b Log β(male) and d' for female judges and the sentence sample



d' scores are varied in the passage sample but, on average, appear to be approximately between the scores of the vowel and sentence samples which indicates that the judges' accuracy was similarly intermediate and confirms the raw percentages of correct gender-identification seen in section 3.1.

Figure 3.4a Log β(male) and d' for male judges and the passage sample

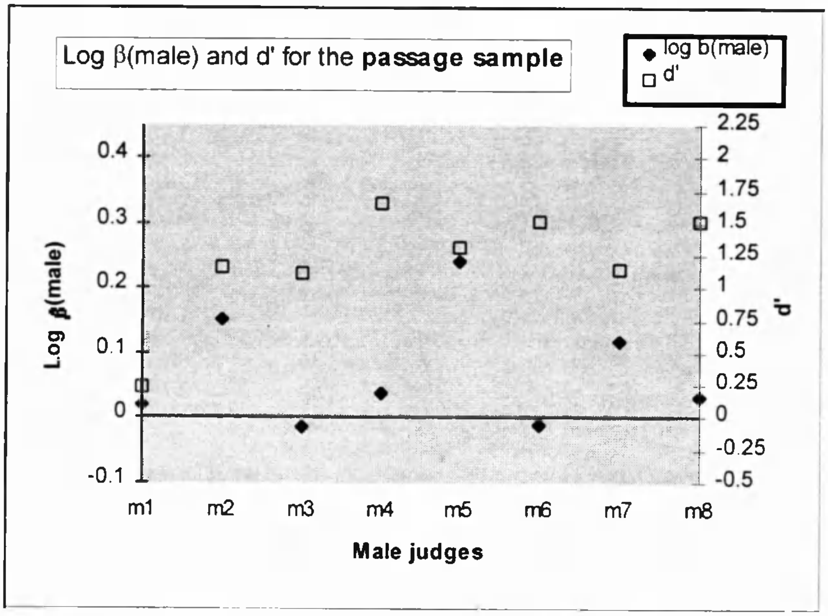
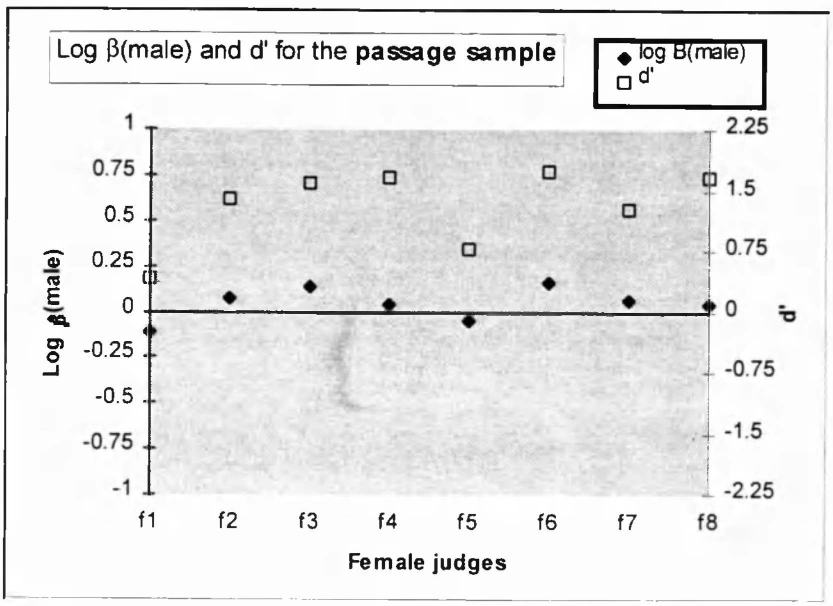


Figure 3.4b Log β(male) and d' for female judges and the passage sample



The means and standard deviations of the β measurement of bias are shown in figure 3.5.

Figure 3.5 Mean and standard deviation values of log β(male) for grouped male and female judges in all speech samples.

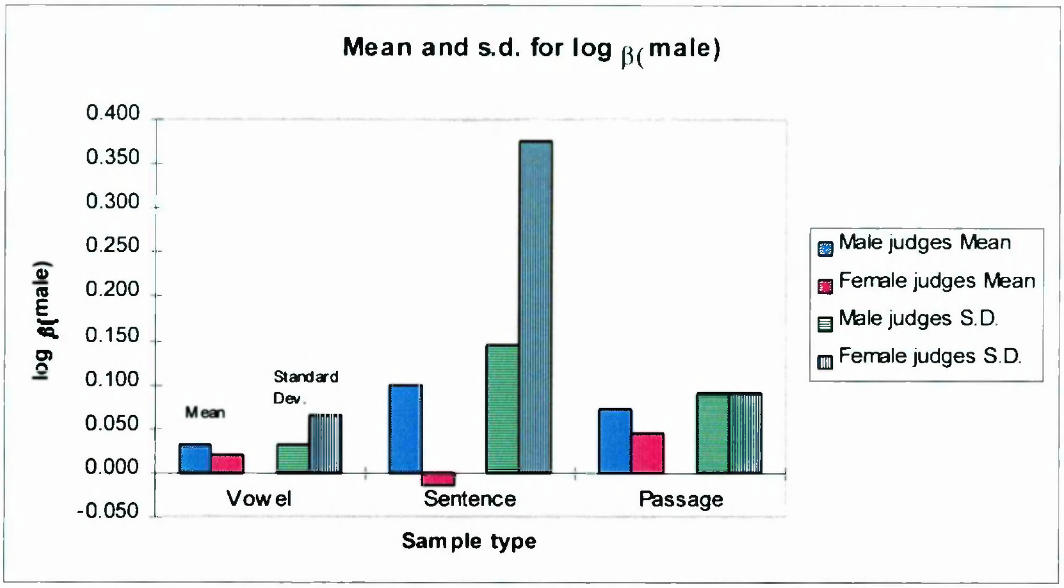
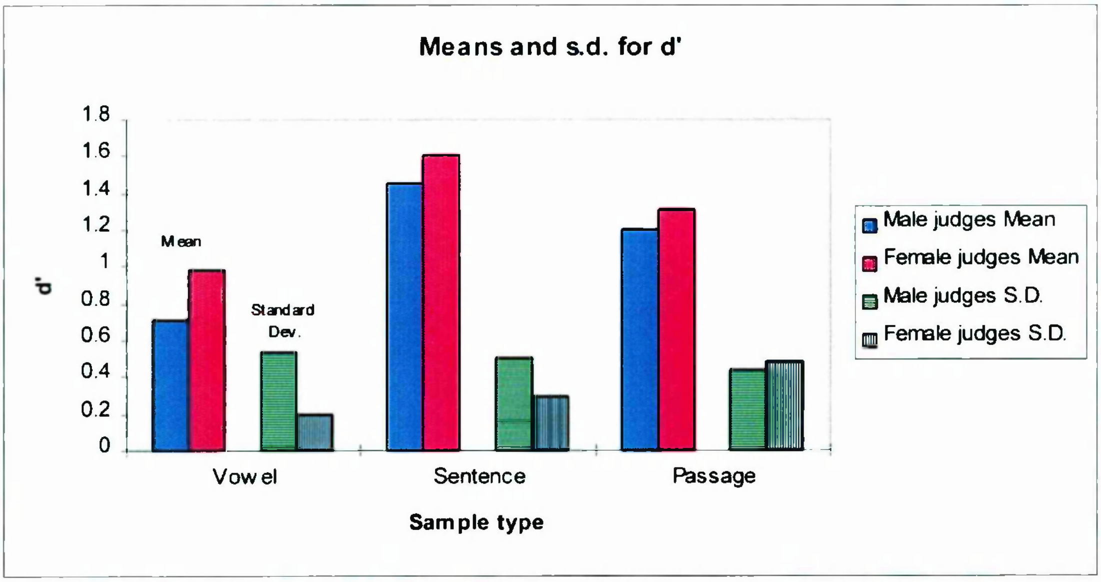


Figure 3.6 shows the mean and standard deviation values for d' (discrimination). We can clearly see that the female judges appear to display a higher level of accuracy in the task than the males and that all judges seem to perform better on the sentence sample than on the passage sample with the vowel sample yielding the lowest identification rates.

Figure 3.6 Mean and standard deviation values of d' for grouped male and female judges in all speech samples.



Figures 3.7 and 3.8 show values of log β(male) and d' respectively for all judges across all sample types. The average scores of each judge are also shown with error bars which demarcate the points +1 S.D. and -1 S.D. around the mean.

Figure 3.7 Values of log β(male) for all judges across all sample types with means and error bars (± 1 S.D.)

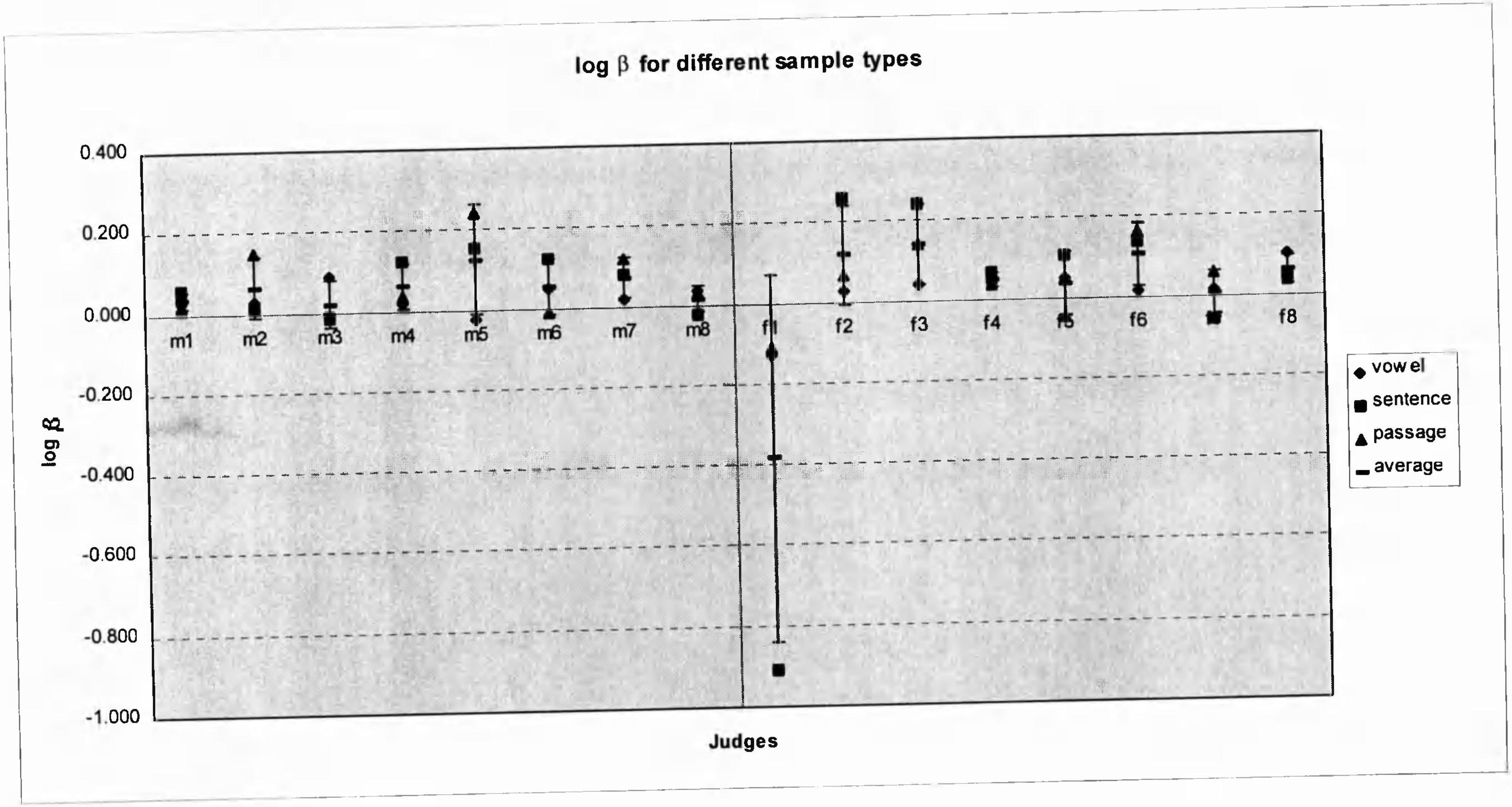
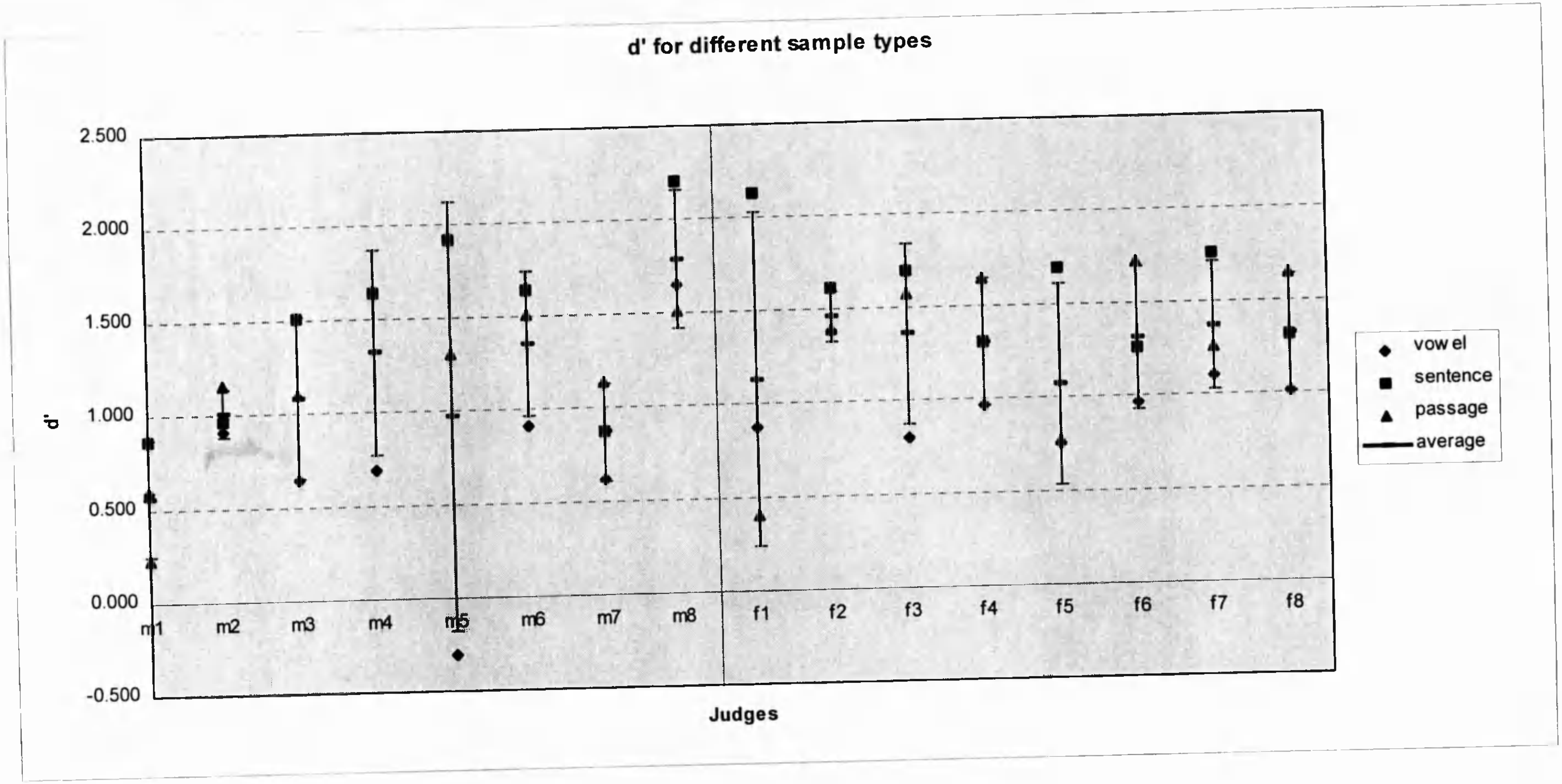


Figure 3.8 Values of d' for all judges across all sample types with means and error bars (± 1 S.D.)



In order to assess the statistical status of any differences which might exist in the data, separate two-factor Analyses of Variance with replication were carried out on the values of log β and d' of the male and female judges. The factors were sex of judge (male or female) and sample type (vowel, sentence or passage). The ANOVA output tables are summarised in tables 3.6 and 3.7

Table 3.6 Results of a two-factor Analysis of Variance between values of log β of male and female judges for all sample types

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Samples	0.00822	2	0.00411	0.133991	0.87497	3.219938
Sex of Judge	0.031482	1	0.031482	1.026389	0.316809	4.07266
Interaction	0.024393	2	0.012196	0.397638	0.674408	3.219938
Within (residual)	1.288237	42	0.030672			
Total	1.352331	47				

Table 3.7 Results of a two-factor Analysis of Variance between values of d' of male and female judges for all sample types

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Samples	3.713449	2	1.856724	10.21523	0.000243	3.219938
Sex of Judge	0.36366	1	0.36366	2.000766	0.164591	4.07266
Interaction	0.060485	2	0.030242	0.166386	0.847275	3.219938
Within (residual)	7.633936	42	0.18176			
Total	11.77153	47				

T-tests were used to compute whether the bias results differed significantly from chance (i.e. the unbiased value of log β=0) . The mean values of all three of the boys' samples differed significantly from chance ($p<0.05$). There were no significant differences between chance and any of the mean values of log β for the girls.

The ANOVA in table 3.7 shows that there was a main effect of sample type in values of d' for male and female judges. In order to identify which of the sample types differed from each other, sign tests were used to assess the differences in the judges' d' scores between each speech sample type (n=16); the results of these tests are shown in tables 3.8a, b and c.

Table 3.8a Results of a sign test between all judges' values of d' for vowel and sentence samples (n=16)

Number of pluses	0
Number of minuses	16
N	16
R_{16}(.05)	**3**
r (test statistic)	**0**
Result	**sig. diff.**

Table 3.8b Results of a sign test for all judges' values of d' between vowel and passage samples (n=16)

Number of pluses	3
Number of minuses	13
N	16
R_{16}(.05)	**3**
r (test statistic)	**3**
Result	**no sig. diff. (marginal)**

Table 3.8c Results of a sign test for all judges' values of d' between sentence and passage samples (n=16)

Number of pluses	10
Number of minuses	6
N	16
R_{16}(.05)	**3**
r (test statistic)	**6**
Result	**no sig. diff.**

Sign tests were also carried out between the d' scores of male and female judges however as the number of pairs in the analysis was so small (n=8) the best confidence level which could be tested would be R(0.1) (i.e. 90% confidence level) and the test was not powerful enough to resolve the differences which appear to exist. Using this test at the 90% confidence level all results showed no significant difference between males and females (c.f. table 3.5).

3.1.2.2 Results following removal of data from biased judge

It is clear from figure 3.9 and from figures 3.2b, 3.3b and 3.4b that one of the judges, f1, is heavily biased to respond 'male' i.e. log $\beta < 1$ (this judge's most extreme β value is marked "outlier" in figure 3.3b). As one of the main purposes of the bias calculations was to identify individuals who might be artificially enhancing or impairing certain aspects of the gender recognition results, and in order to present a more reliable picture of the pattern of judge behaviour, the gender recognition results have been recalculated with the data from judge f1 removed from the analysis. These results are shown in tables 3.9 and 3.10.

Table 3.9 Mean correct recognition rates for male and female children by male and female listeners across vowel, sentence and spontaneous speech conditions. Data from judge f1 has been removed (Brackets indicate percentage change from raw results)

VOWELS

	Male Judges	Female Judges	AVERAGE
Male children	59%	65% (-2)	61.96% (-0.81)
Female children	68%	72% (+2)	70.22% (+1.18)
AVERAGE	63.36%	68.82% (+0.37)	**66.08% (+0.17)**

SENTENCES

	Male Judges	Female Judges	AVERAGE
Male children	71%	72% (-3)	71.49% (-1.51)
Female children	80%	82% (+3)	81.17% (+1.81)
AVERAGE	75.43%	77.22% (+0.19)	**76.33% (+0.1)**

SPONTANEOUS SPEECH

	Male Judges	Female Judges	AVERAGE
Male children	67%	73% (-1)	69.93% (-0.45)
Female children	77%	78% (+5)	77.72% (+2.87)
AVERAGE	71.81%	75.85% (+2.42)	**73.83% (+1.21)**

Table 3.10 Total number of 'male' and 'female' responses made by male and female listeners to male and female children across vowel, sentence and spontaneous speech conditions exluding data from judge f1.

VOWELS

	Male Judges		Female Judges		TOTALS	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
Male children	216	152	210	112	426	264
Female children	110	234	83	218	193	452
TOTALS	326	386	293	330	619	716
Percentage of total responses	45.8%	54.2%	47%	53%	46.4%	53.6%

SENTENCES

	Male Judges		Female Judges		TOTALS	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
Male children	261	107	232	90	493	197
Female children	69	275	53	248	122	523
TOTALS	330	382	285	338	615	720
Percentage of total responses	46.3%	53.7%	45.7%	54.3%	46.1%	53.9%

SPONTANEOUS SPEECH

	Male Judges		Female Judges		TOTALS	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
Male children	245	123	236	86	481	209
Female children	79	265	65	236	144	501
TOTALS	324	388	301	322	625	710
Percentage of total responses	45.5%	54.5%	48.3%	51.7%	46.8%	53.2%

Tables 3.11 and 3.12 show the results of analyses of variance on the correct recognition data with judge f1 removed from the comparisons. The recalculated comparison of accuracy of gender identification indicates that a new pattern of results holds from the raw data - i.e. that the difference between the judgements of boys and girls is now significant, with the girls being better identified than the boys ($p=0.012$). Overall, the difference between male and female judges' accuracy in gender identification measured across all children, was not significant in the 'by judges' analysis ($p=0.161$).

The Scheffé tests performed on the means of the material type variable in both the 'by children' and 'by judges' analyses yielded similar results to the corresponding tests with the raw data (see appendix 4). In the 'by children' analysis, the differences between the identification rates of the sample types were not significant, however the result suggests that a significant difference would be most likely seen between the vowel and sentence samples - a fact confirmed by the relative sizes of the means. In the 'by judges' analysis, the difference between vowel and sentence samples was significant ($p<0.05$) and the results indicate that the passage sample was probably (but not definitely) identified at a better rate than the vowel sample.

Table 3.11 Results of a 3 - Way Analysis of Variance with sex of child as the grouping (between subjects) variable and sex of judge and material type as repeated measures. Data from judge f1 has been removed.

Source	Sum of Squares	D.F.	Mean Square	F	Tail Prob.
Main Effects					
Sex of child	0.981	1	0.981	6.58	0.012
ERROR	12.963	87	0.149		
Sex of judge	0.189	1	0.189	10.55	0.002
ERROR	1.555	87	0.018		
Material type	1.013	2	0.506	3.69	0.027
ERROR	23.882	174	0.137		
Interactions					
Judge x Child	0.014	1	0.014	0.78	0.379
Material x Child	0.009	2	0.004	0.03	0.969
Judge x Material	0.031	2	0.015	0.77	0.463
Judge x Material x Child	0.025	2	0.012	0.62	0.537
ERROR	3.443	174	0.020		

Note the lack of interaction between sex of child and sex of judge. This contrasts with the results of the analysis using all judges in which there was a strong interaction (see fig. 3.1). The difference between these situations is due to the fact that, because judge f1 was biased to respond 'boy' to all children, she yielded an artificially high rate of correct identification scores for the boys and an artificially low rate for girls. This influenced the overall mean scores and effectively boosted the apparent correct identification rates of the boys and lowered the apparent correct identification rates of the girls. This is the reason for the 'cross-over' effect seen in fig. 3.1. When judge f1 is removed from the data, we see what is probably a truer pattern of results (fig. 3.9 below) i.e. that female judges perform more accurately than male judges and that girls are better identified than boys by all judges. Female judges are significantly better than male judges at identifying boys (p=0.01) but the analogous test using girls narrowly fails to achieve a significant results (p=0.07)

Figure 3.9 Chart showing lack of interaction between sex of child and sex of judge data when responses from judge f1 are removed

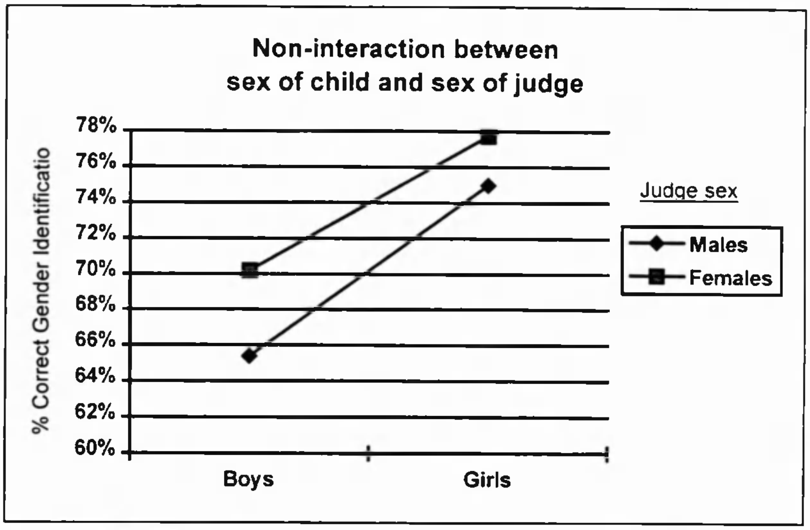


Table 3.12 Results of a 3 - Way Analysis of Variance with sex of judge as the grouping (between subjects) variable and sex of child and material type as repeated measures. Data from judge f1 has been removed.

Source	Sum of Squares	D.F.	Mean Square	F	Tail Prob.
Main Effects					
Sex of child	0.165	1	0.165	37.55	0.0001
ERROR	0.057	13	0.004		
Sex of judge	0.032	1	0.032	2.21	0.161
ERROR	0.187	13	0.014		
Material type	0.170	2	0.085	12.56	0.0002
ERROR	0.176	26	0.007		
Interactions					
Judge x Child	0.002	1	0.002	0.53	0.48
Material x Child	0.001	2	0.001	0.24	0.790
Judge x Material	0.005	2	0.003	0.38	0.689
Judge x Material x Child	0.004	2	0.002	0.69	0.510
ERROR	0.079	26	0.003		

The mean log β and d' results have also been computed with f1 removed from the analysis. These results are shown in figures 3.10 and 3.11 respectively. Note that with f1 removed, the pattern of log β results is now more regular with all judges showing a bias towards responding 'girl' in all sample types. The very large standard deviation of the female judges in the sentence sample which was evident when all judges were included has now been normalised and it is clear that judge fl was responsible for a large part of the earlier variance. The newly calculated d' results, seen in figure 3.11, are largely unchanged from the original results (figure 3.6) as judge fl's discrimination scores were reasonably normal.

Figure 3.10 Mean and standard deviation values of log β(male) for grouped male and female judges excluding judge fl (outlier) in all speech samples.

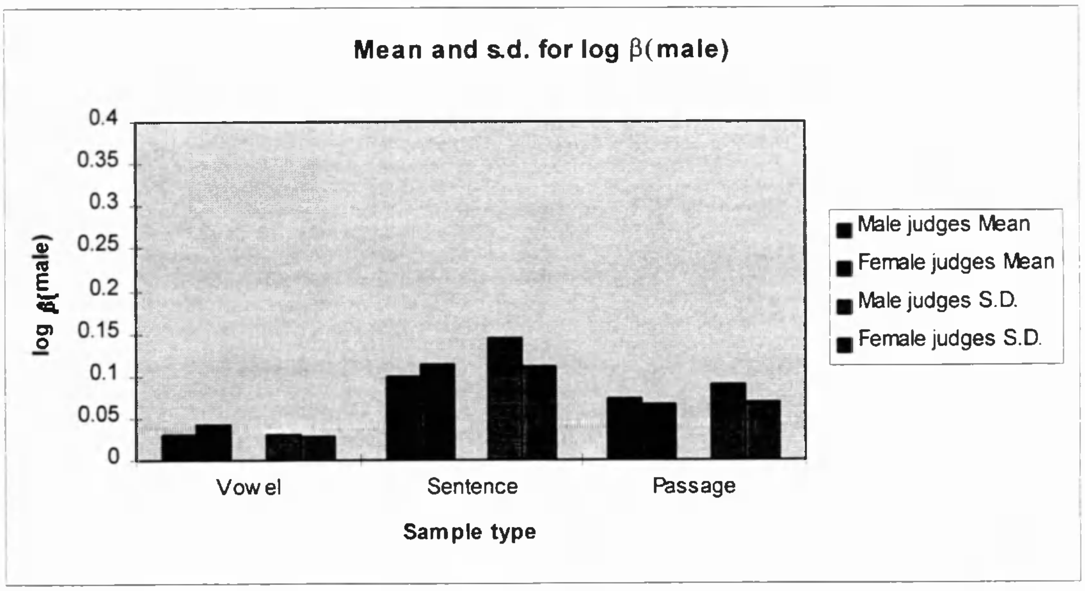
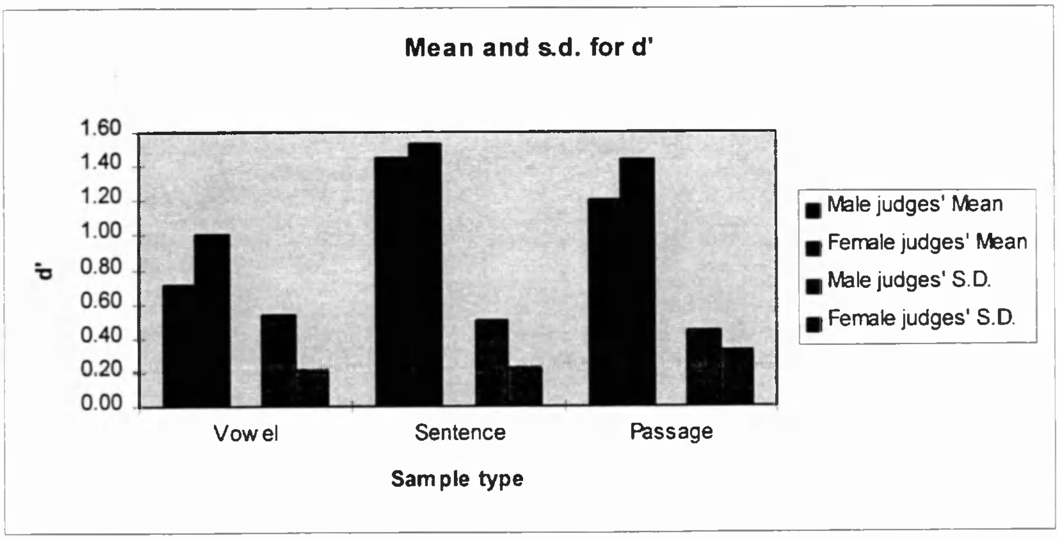


Figure 3.11 Mean and standard deviation values of d' for grouped male and female judges excluding judge fl (outlier) in all speech samples.



In order to assess the statistical status of any differences which might emerge in the bias and discrimination data after removal of judge f1's responses, separate two-factor Analyses of Variance with replication were carried out on the values of log β and d' of the male and female judges as before. The factors were sex of judge (male or female) and sample type (vowel, sentence or passage). The ANOVAs indicated that there were no differences between log β or d' according to the sex of judge or material type with the exception of the comparison of d' values across different samples. This result was significant which indicates that judges performed with significantly different degrees of accuracy to the three different sample types. This further confirms the observation that the sentences generally yielded better rates of correct gender identification than the other sample types. The ANOVA output tables are summarised in appendix 6.

T-tests were used to compute whether the bias results, without data from judge f1, differed significantly from chance (i.e. the unbiased value of log $\beta=0$) . The mean log β values of the boys' vowel and passage samples differed significantly from chance ($p<0.05$), the mean values of the sentence sample narrowly failed to achieve significance ($p=0.06$). The mean log β values of all three samples of the girls (which had failed to achieve a significant difference when all judges were included) were significantly different from chance when judge f1 was removed ($p<0.05$).

<u>Summary of Perceptual Results</u>

- The gender of prepubertal children was identified at better than chance rates
- Girls were identified significantly better than boys
- Female judges were significantly better than male judges at identifying boys but overall there was no statistically significant difference in the correct recognition ability of male and female judges.
- There was a significant difference in the correct identification rates across sample types. The sentence sample was best identified, the vowel sample was the worst identified.
- Listeners were slightly biased to respond female

3.2 ACOUSTIC ANALYSIS RESULTS

3.2.1 Vowels

3.2.1.1 Vowel Formant Information

The mean values of the first three formant frequencies (F1-3), bandwidths (B1-3) and amplitudes (A1-3) of the isolated vowel productions of all of the children are shown in tables 3.13-3.15 below. T-tests were carried out to assess the differences in formant measures between the sexes. These show significant differences between the sexes of mean values of F1 and F2 of [a], and F2 of [i]. A1 and A2 of [a] and A3 of [o] also showed significant sex-differences and there were a number of marginal failures of significance. It would appear therefore that /a/ is the vowel which best reveals sex-specific acoustic differences. This is most likely due to the fact that the production of /a/ involves a relatively open, unconstricted vocal tract and therefore would best reflect any differences in the sizes of the resonating cavities. It is also a vowel which is robust enough to allow a number of voluntary articulations (such as voluntary, sex-specific modifications perhaps) without compromising the phonemic identity of the sound.

Table 3.13 Formant frequencies, bandwidths and amplitudes of all children's productions of vowel [a] arranged by sex. Standard deviations are in parenthesis. Results of a t-test between mean formant values for boys and girls are also shown.

	Boys	Girls	t	*p*
F1 (Hz)	1353.799 (149.268)	1447.312 (162.631)	2.510	0.015
B1 (Hz)	510.829 (131.648)	540.933 (160.219)	-0.856	0.395
A1 (dB)	68.442 (6.343)	64.513 (6.228)	2.629	0.011
F2 (Hz)	2145.498 (196.935)	2269.381 (178.830)	2.777	0.007
B2 (Hz)	397.954 (130.712)	418.457 (147.139)	-0.617	0.540
A2 (dB)	67.169 (6.488)	63.738 (6.122)	2.290	0.025
F3 (Hz)	3774.029 (274.199)	3853.139 (248.363)	1.275	0.206
B3 (Hz)	471.298 (228.858)	502.164 (272.568)	-0.512	0.610
A3 (dB)	59.817 (5.218)	57.498 (4.965)	1.917	0.059

Table 3.14 Formant frequencies, bandwidths and amplitudes of all children's productions of vowel [i] arranged by sex. Standard deviations are in parenthesis. Results of a t-test between mean formant values for boys and girls are also shown.

	Boys	Girls	t	*p*
F1 (Hz)	391.671 (34.151)	410.515 (63.619)	1.579	0.120
B1 (Hz)	307.066 (100.263)	292.228 (79.696)	0.712	0.479
A1 (dB)	60.733 (6.190)	58.614 (4.661)	1.683	0.097
F2 (Hz)	3416.636 (252.501)	3592.435 (306.799)	2.496	0.016
B2 (Hz)	476.790 (147.465)	459.589 (183.254)	0.260	0.796
A2 (dB)	54.042 (6.634)	53.149 (8.993)	0.306	0.760
F3 (Hz)	3836.019 (265.573)	3934.585 (209.008)	1.793	0.077
B3 (Hz)	509.075 (281.208)	431.871 (287.080)	1.175	0.244
A3 (dB)	54.921 (6.266)	53.563 (6.974)	0.885	0.379

Table 3.15 Formant frequencies, bandwidths and amplitudes of all children's productions of vowel [o] arranged by sex. Standard deviations are in parenthesis. Results of a t-test between mean formant values for boys and girls are also shown.

	Boys	Girls	t	*p*
F1 (Hz)	572.632 (53.838)	592.697 (58.072)	1.472	0.145
B1 (Hz)	256.175 (82.032)	270.240 (85.449)	-0.746	0.458
A1 (dB)	66.011 (4.787)	63.903 (5.141)	1.826	0.072
F2 (Hz)	2322.747 (120.674)	2398.659 (214.938)	1.957	0.055
B2 (Hz)	181.105 (181.681)	241.313 (200.664)	-1.463	0.148
A2 (dB)	33.514 (4.157)	32.244 (5.332)	1.048	0.298
F3 (Hz)	3851.771 (314.749)	3846.543 (299.995)	-0.20	0.843
B3 (Hz)	808.856 (267.949)	924.974 (286.793)	-1.734	0.087
A3 (dB)	38.101 (6.342)	35.086 (5.697)	2.164	0.034

The formant scaling factors (k-factors) of the boys and girls were calculated according to Fant's (1966) formula,

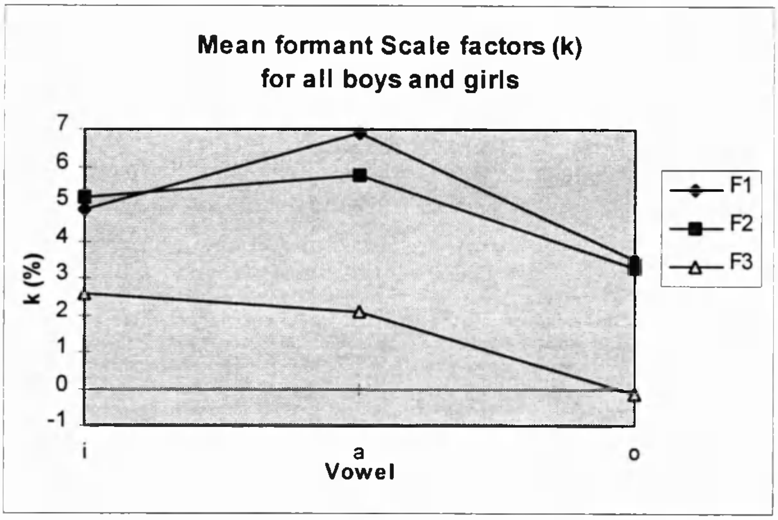
$$k_n = \left[\frac{\mathrm{F}n \text{ of female}}{\mathrm{F}n \text{ of male}} - 1\right] \times 100$$

These results are presented in table 3.16 and figure 3.12.

Table 3.16 Formant scaling factors (k factors) for mean formant frequency values of boys and girls

	i	a	o		MEAN
F1	4.811	6.907	3.504		**5.074**
F2	5.145	5.774	3.268		**4.729**
F3	2.569	2.096	-0.136		**1.510**
MEAN	**4.175**	**4.926**	**2.212**		**3.771**

Figure 3.12 Chart of formant scaling factors (k factors) for mean formant frequency values of boys' and girls' productions of three test vowels.



Clearly, /a/ yielded the largest sex-differences, as expected, and /o/ yielded the smallest. We can also see that the sex-differences involved in F1 and F2 were much larger than those of F3.

A multiple regression analysis using formant frequency values and maleness scores was carried out and the results are shown below in table 3.17. The reader will recall that the multiple regression analysis is concerned with the construction of a linear equation which predicts the values of a target (dependent) variable from the knowledge of specified values of a set of regressors (independent variables). In this case the dependent variable is the maleness score of each child and the independent variables are the formant frequency values of each child.

Table 3.17 Simultaneous regression output table of maleness scores upon nine formant frequency regressors

Multiple R	.734
R Square	**.539**
Adjusted R Square	.414
Standard Error	3.864

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	9	577.046	64.116
Residual	33	492.815	14.934

F = 4.29337 Signif. F = .0009

Variables in the Equation

Variable	B	SE B	Beta	T	Sig.T
F1 of /a/	.012499	.004115	.451514	3.037	.0046
F1 of /i/	-.016558	.016036	-.130238	-1.033	.3093
F1 of /o/	-.004479	.011791	-.052663	-.380	.7065
F2 of /a/	-.010384	.004592	-.346256	-2.262	.0304
F2 of /i/	-.007478	.002681	-.364201	-2.789	.0087
F2 of /o/	-.001820	.003187	-.079184	-.571	.5719
F3 of /a/	-.005604	.002178	-.340112	-2.573	.0148
F3 of /i/	1.67463E-04	.002861	.007407	.059	.9537
F3 of /o/	.001857	.002194	.118181	.846	.4035

The shaded cells in the table of variables above represent those formant frequencies which make a significant contribution to accounting for the variance in the dependent variable. The reader will note that these variables are F1, F2 and F3 of /a/ and F2 of /i/.

Similar regressions were carried out on the formant amplitude and bandwidth data. In the amplitude analysis, only A2 of [a] was significantly involved in accounting for the overall variance (Sig. T=0.0013).

The R square value of the regression was 0.345. No individual measures of formant bandwidth significantly accounted for the variance. R square for the bandwidth data was 0.294.

Each child received either a 'male' or 'female' judgement from each judge. There were 16 judges, therefore a maleness score could be calculated for each child based on the number of 'male' judgements assigned to that particular child (max. 16). Those children who scored 15 or 16 were included in the *most male-like* group of the analysis while those children who scored 0 or 1 were included in the *least male-like* group. This analysis therefore investigates the differences between groups of children who were perceived to be boys and girls without reference to their actual biological gender.[16]

The formant results of the previous analysis are arranged according to the children who were judged to be most male-like and least male-like (i.e. most boy-like and girl-like respectively). A statistical comparison was made between the two groups using t-tests. These results are shown in tables 3.18-3.20 below. Notice that in this analysis, F2 of /a/ and F1 of /o/ both show significant differences between the perceived-gender groups. Unlike the previous analysis, certain bandwidth comparisons yielded significance (B1 of /a/ and B2 of /o/). Finally, seven of the nine amplitude comparisons achieved statistical significance. See section 4.2.1.1 for a detailed discussion of these results.

[16] See Appendix 8 for a list of most and least male-like children

Table 3.18 Formant frequencies, bandwidths and amplitudes of most male-like and least male-like children's productions of vowel [a] arranged by group. Standard deviations are in parenthesis. Results of a t-test between mean formant frequency values for the two groups are also shown.

	Most male-like	Least male-like	t	*p*
F1 (Hz)	1168.042 (324.182)	1125.88 (409.735)	-0.286	0.389
B1 (Hz)	454.250 (101.937)	563.231 (188.966)	1.864	0.038
A1 (dB)	67.820 (7.890)	58.605 (6.297)	-3.094	0.003
F2 (Hz)	2009.956 (354.712)	2263.848 (213.784)	2.031	0.032
B2 (Hz)	501.317 (202.809)	534.927 (251.384)	0.368	0.358
A2 (dB)	67.542 (5.711)	61.352 (6.485)	-2.514	0.010
F3 (Hz)	3794.284 (230.282)	3875.863 (195.312)	0.921	0.185
B3 (Hz)	326.187 (241.697)	421.176 (304.101)	0.867	0.198
A3 (dB)	58.190 (6.108)	56.334 (3.528)	-0.869	0.200

Table 3.19 Formant frequencies, bandwidths and amplitudes of most male-like and least male-like children's productions of vowel [i] arranged by group. Standard deviations are in parenthesis. Results of a t-test between mean formant frequency values for the two groups are also shown.

	Most male-like	Least male-like	t	*p*
F1 (Hz)	397.602 (98.655)	410.070 (70.157)	0.346	0.367
B1 (Hz)	291.805 (151.101)	306.237 (106.305)	0.257	0.400
A1 (dB)	64.823 (3.206)	57.764 (6.048)	-3.792	0.001
F2 (Hz)	3214.413 (522.598)	3311.276 (620.330)	0.421	0.339
B2 (Hz)	463.503 (80.243)	342.521 (179.315)	-2.291	0.016
A2 (dB)	56.326 (3.686)	47.572 (9.793)	-3.144	0.003
F3 (Hz)	3596.004 (1022.77)	3559.421 (1316.14)	-0.078	0.469
B3 (Hz)	578.272 (250.800)	520.006 (254.431)	-0.549	0.295
A3 (dB)	57.071 (5.367)	52.609 (4.973)	-2.096	0.025

Table 3.20 Formant frequencies, bandwidths and amplitudes of most male-like and least male-like children's productions of vowel [o] arranged by group. Standard deviations are in parenthesis. Results of a t-test between mean formant frequency values for the two groups are also shown.

	Most male-like	**Least male-like**	**t**	***p***
F1 (Hz)	507.143 (71.191)	616.625 (22.012)	4.243	0.001
B1 (Hz)	277.42 (92.499)	293.813 (84.430)	0.417	0.342
A1 (dB)	66.101 (4.178)	62.655 (4.413)	-1.845	0.042
F2 (Hz)	2280.427 (62.678)	2363.024 (382.793)	0.815	0.214
B2 (Hz)	111.548 (89.042)	289.144 (153.133)	3.514	0.001
A2 (dB)	34.294 (3.146)	33.888 (6.212)	-0.208	0.419
F3 (Hz)	3886.425 (284.101)	3859.934 (345.065)	-0.197	0.423
B3 (Hz)	860.634 (285.145)	879.390 (265.533)	0.154	0.440
A3 (dB)	41.531 (5.438)	33.147 (3.766)	-3.891	0.001

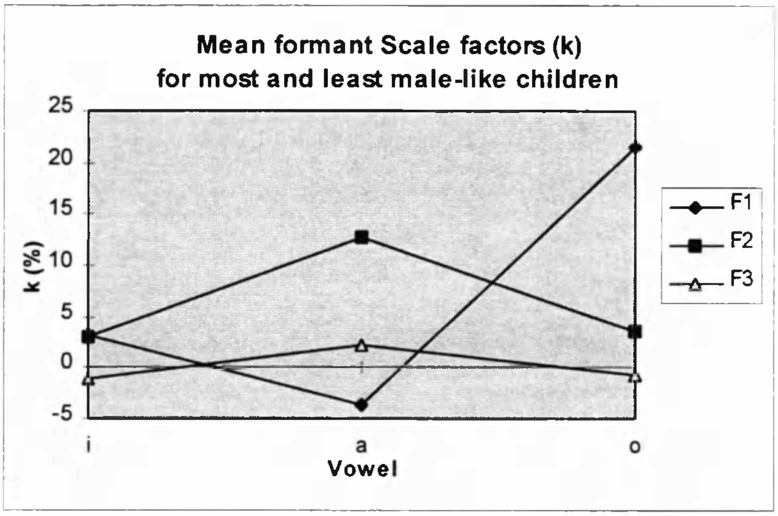
The mean values of the first three formant frequencies, bandwidths and amplitudes for most and least male-like children are also displayed graphically in appendix 7.

The formant scaling factors of the most and least male-like children are shown in table 3.21 and figure 3.13 below.

Table 3.21 Formant scaling factors (k factors) for mean formant frequency values of most and least male-like children

	i	a	o		MEAN
F1	3.136	-3.610	21.588		**7.038**
F2	3.013	12.632	3.622		**6.422**
F3	-1.017	2.150	-0.682		**0.150**
MEAN	**1.711**	**3.724**	**8.176**		**4.537**

Figure 3.13 Chart of formant scaling factors (k factors) for mean formant frequency values of most and least male-like children's productions of three test vowels.



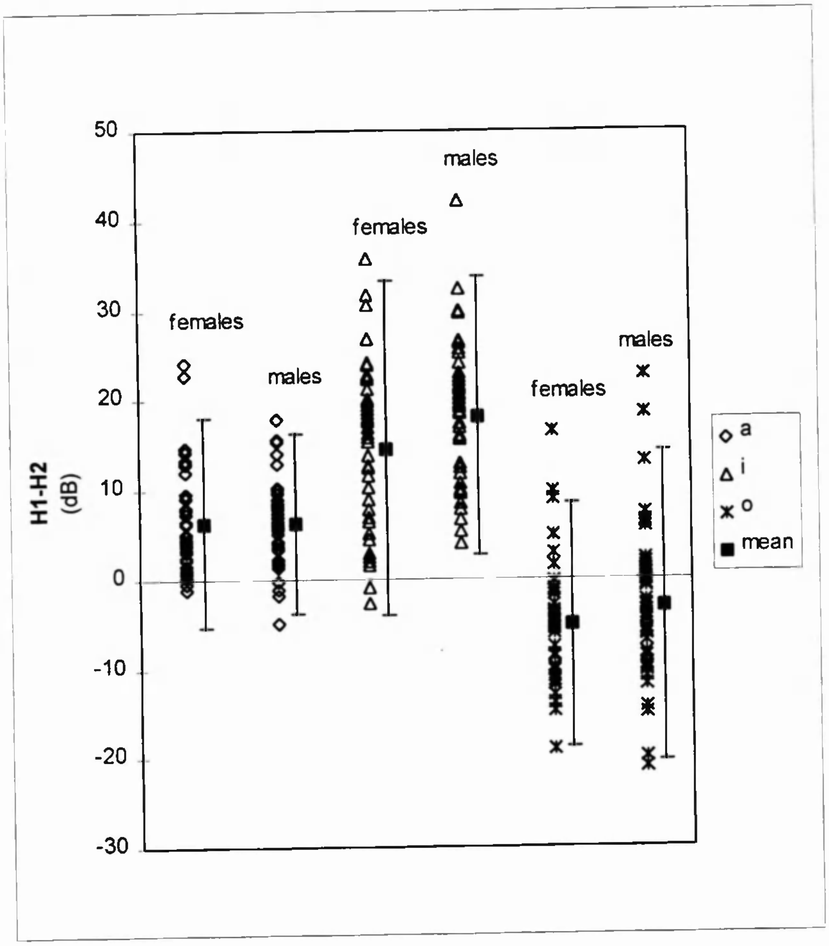
From tables 3.18-3.20 it would appear that the formant amplitudes are involved to a greater degree than in the earlier 'boys vs girls' analysis. Seven of the nine comparisons show significant results. Therefore a multiple regression analysis was carried out on the amplitude data in order to assess how much of the variance in the maleness scores could be accounted for by differences in formant amplitude. The F-ratio of the analysis of variance which accompanies the regression was non-significant ($F=1.738$; $p=0.183$), therefore the regressors were not able to account for enough of the variance to make the analysis meaningful.

3.2.1.2 Breathiness Measures

<u>H1-H2 measure</u>

The first measure of breathiness reflects the difference in amplitude of the first two harmonics and will be termed 'H1-H2'. Figure 3.14 shows the H1-H2 values of each individual child's production of each vowel with means and error bars. Table 3.22 gives group mean and standard deviations for boys and girls expressed by vowel.[17]

Figure 3.14 H1-H2 values for all children across all vowels (/a, i, o/)



17 See section 4.2.1.2 for a discussion of the validity of the breathiness results.

Table 3.22 Mean and standard deviation values of the breathiness parameter H1-H2 (dB).

	Vowel		
	a	i	o
boys mean	6.14	18.12	-3.09
girls mean	6.31	14.51	-5.23
boys s.d.	*5.05*	*7.79*	*8.65*
girls s.d.	*5.82*	*9.34*	*6.79*

The H1-H2 values of boys and girls were analysed using t-tests in order to determine if there was a significant difference between the sexes. The results of the 3 t-tests are summarised in table 3.23 a-c below. Note that differences in breathiness in /i/ approached significance, all other comparisons were non-significant.

Table 3.23a, 3.23b and 3.23c T-tests comparing boys' and girls' values of the breathiness parameter H1-H2.

Vowel is [a]		
	Females	*Males*
Mean	6.310465	6.140222
Variance	33.83404	25.46551
Observations	43	45
Hypothesized Mean Difference	0	
df	83	
t Stat	0.146373	
P(T<=t) one-tail	0.441991	
t Critical one-tail	1.66342	
P(T<=t) two-tail	0.883982	**N.S.**
t Critical two-tail	1.98896	

Vowel is [i]		
	Females	*Males*
Mean	14.51395	18.12022
Variance	87.171	60.60774
Observations	43	46
Hypothesized Mean Difference	0	
df	82	
t Stat	-1.97185	
P(T<=t) one-tail	0.025999	
t Critical one-tail	1.663648	
P(T<=t) two-tail	0.051999	1>p>0.05
t Critical two-tail	1.98932	

Vowel is [o]		
	Females	*Males*
Mean	-5.23048	-3.08804
Variance	46.15419	74.85717
Observations	42	46
Hypothesized Mean Difference	0	
df	84	
t Stat	-1.29755	
P(T<=t) one-tail	0.098996	
t Critical one-tail	1.663198	
P(T<=t) two-tail	0.197993	**N.S.**
t Critical two-tail	1.98861	

As can be seen from figure 3.14, the majority of the data are reasonably tightly distributed but there are a number of outliers in the H1-H2 data points. These outliers will reduce the accuracy of the mean as a reflection of true central tendency. An outlier is defined here as any point which is more than 2 S.D.s from the mean. The H1-H2 values have therefore been recalculated with those points which conform to the above definition of outlier removed. This data is shown in figure 3.15 and table 3.24.

Figure 3.15 H1-H2 values for all children across all vowels (/a, i, o/) with outliers removed.

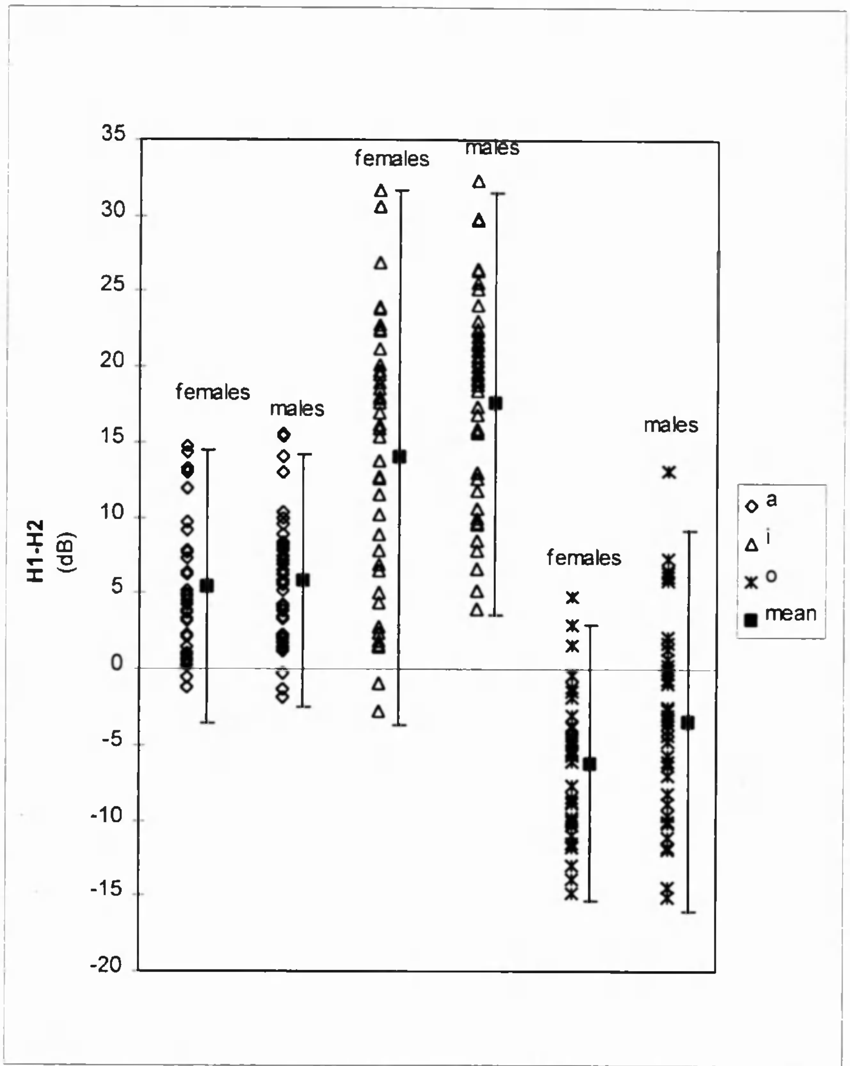


Table 3.24 Mean and standard deviation values of the breathiness parameter H1-H2 (dB) with outlier data removed.

	Vowel		
	a	i	o
boys mean	5.85	17.59	-3.39
girls mean	5.48	14.01	-6.20
boys s.d.	*4.19*	*6.97*	*6.29*
girls s.d.	*4.50*	*8.84*	*4.56*
number of outliers removed:			
boys	**3**	**1**	**4**
girls	**2**	**1**	**4**

T-tests comparing H1-H2 values for the children's data without outliers are summarised in table 3.25 a-c. H1-H2 differed significantly by sex in the vowels /i/ and /o/ but not in /a/.

Table 3.25a, 3.25b and 3.25c t-tests comparing boys' and girls' values of the breathiness parameter H1-H2 without outliers

Vowel is [a]		
	Females	*Males*
Mean	5.478	5.849
Variance	20.24	17.58
Observations	41	42
Hypothesized Mean Difference	0	
df	80	
t Stat	-0.389	
P(T<=t) one-tail	0.349	
t Critical one-tail	1.664	
P(T<=t) two-tail	0.699	**N.S.**
t Critical two-tail	1.99	

Vowel is [i]		
	Females	*Males*
Mean	14.0119	17.58578
Variance	78.19449	48.54779
Observations	42	45
Hypothesized Mean Difference	0	
df	78	
t Stat	-2.08411	
P(T<=t) one-tail	0.020212	
t Critical one-tail	1.664625	
P(T<=t) two-tail	0.040424	**p<0.05**
t Critical two-tail	1.990848	

Vowel is [o]		
	Females	*Males*
Mean	-6.20263	-3.38809
Variance	20.77145	39.52346
Observations	38	42
Hypothesized Mean Difference	0	
df	75	
t Stat	-2.30757	
P(T<=t) one-tail	0.011891	
t Critical one-tail	1.665426	
P(T<=t) two-tail	0.023783	**p<0.05**
t Critical two-tail	1.992103	

H1-F1 measure

The second breathiness measure reflects the difference in amplitudes between the first harmonic (fundamental) and the first formant and is labelled 'H1-F1'. Figure 3.16 shows H1-F1 values of each individual child's production of each vowel with means and error bars - outliers have been removed from this data. Table 3.26 gives group mean and standard deviations for boys and girls expressed by vowel.

Figure 3.16 H1-F1 values for all children across all vowels (/a, i, o/) with outliers removed

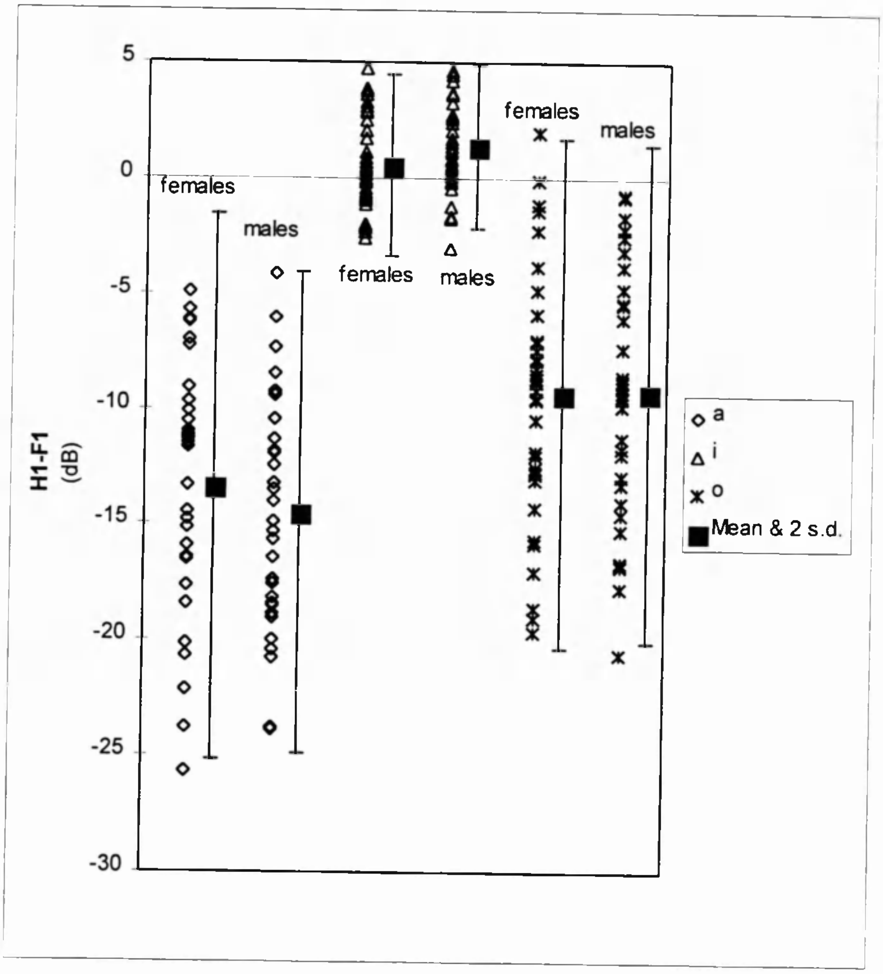


Table 3.26 Mean and standard deviation values of the breathiness parameter H1-F1 (dB) with outlier data removed

	Vowel		
	a	i	o
boy's mean	-14.49	1.37	-9.26
girl's mean	-13.39	0.55	-9.32
boy's s.d.	*5.21*	*1.77*	*5.38*
girl's s.d.	*5.89*	*1.95*	*5.52*

The H1-F1 values of boys and girls were analysed using t-tests in order to determine if there was a significant difference between the sexes. The results of the 3 t-tests are summarised in tables 3.27a-c below. None of the comparisons showed significant differences although the test for /i/ indicated a marginal failure of significance.

Table 3.27a, 3.27b and 3.27c t-tests comparing boys' and girls' values of the breathiness parameter H1-F1 without outliers

t-Test: Vowel [a]		
	Females	*Males*
Mean	-13.3909	-14.4888
Variance	34.73199	27.19238
Observations	32	35
Hypothesized Mean Difference	0	
df	62	
t Stat	0.804522	
P(T<=t) one-tail	0.212084	
t Critical one-tail	1.669804	
P(T<=t) two-tail	0.424169	**N.S.**
t Critical two-tail	1.998969	

t-Test: Vowel [i]		
	Females	*Males*
Mean	0.547022	1.366252
Variance	3.803104	3.116826
Observations	40	37
Hypothesized Mean Difference	0	
df	75	
t Stat	-1.93462	
P(T<=t) one-tail	0.028404	
t Critical one-tail	1.665426	
P(T<=t) two-tail	0.056808	1>p>0.05
t Critical two-tail	1.992103	

t-Test: Vowel [o]		
	Females	*Males*
Mean	-9.32308	-9.26462
Variance	30.41932	28.9241
Observations	36	35
Hypothesized Diff.	0	
df	69	
t Stat	-0.04522	
P(T<=t) one-tail	0.482031	
t Critical one-tail	1.667238	
P(T<=t) two-tail	0.964062	**N.S.**
t Critical two-tail	1.994945	

The H1-F1 values of the most and least male-like children were analysed using t-tests in order to determine if there was a significant difference between the children whom listeners identified perceptually as being most typically male or female. The results of the 3 t-tests are summarised in tables 3.28a-c below. Once again there were no significant differences between the breathiness values of most and least male-like children for any of the vowels (vowel /o/ showed a marginal failure).

Table 3.28a, 3.28b and 3.28c t-tests comparing values of the breathiness parameter H1-F1 of children judged to be most and least male-like

[a]	*least male-like*	*most male-like*
Mean	-10.344	-13.348
Variance	18.147	35.085
Observations	9	7
Hypothesized Mean Difference	0	
df	11	
t Stat	1.133	
P(T<=t) one-tail	0.141	
t Critical one-tail	1.796	
P(T<=t) two-tail	0.281	**N.S.**
t Critical two-tail	2.201	

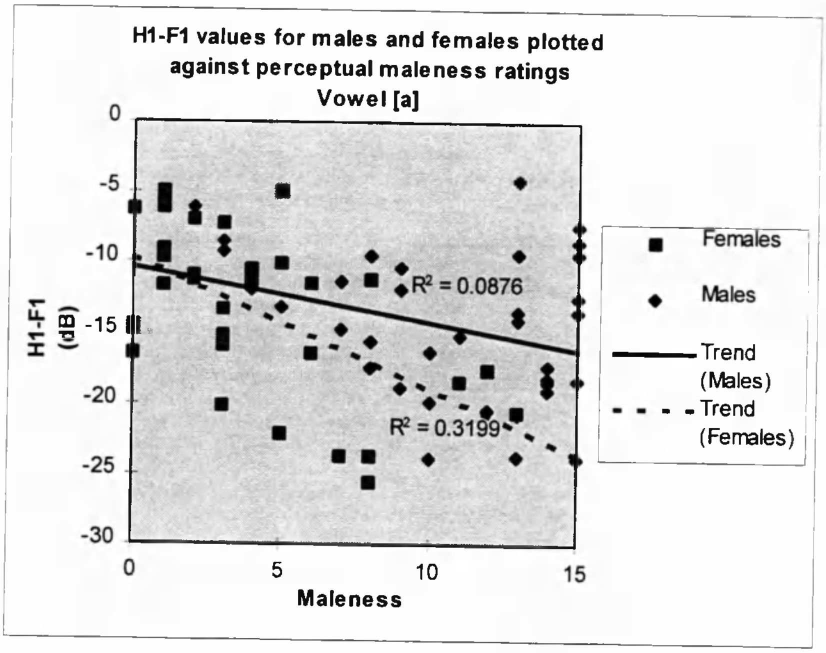
[i]	*least male-like*	*most male-like*
Mean	1.079	1.301
Variance	4.666	4.153
Observations	14	9
Hypothesized Mean Difference	0	
df	18	
t Stat	-0.249	
P(T<=t) one-tail	0.403	
t Critical one-tail	1.734	
P(T<=t) two-tail	0.806	**N.S.**
t Critical two-tail	2.101	

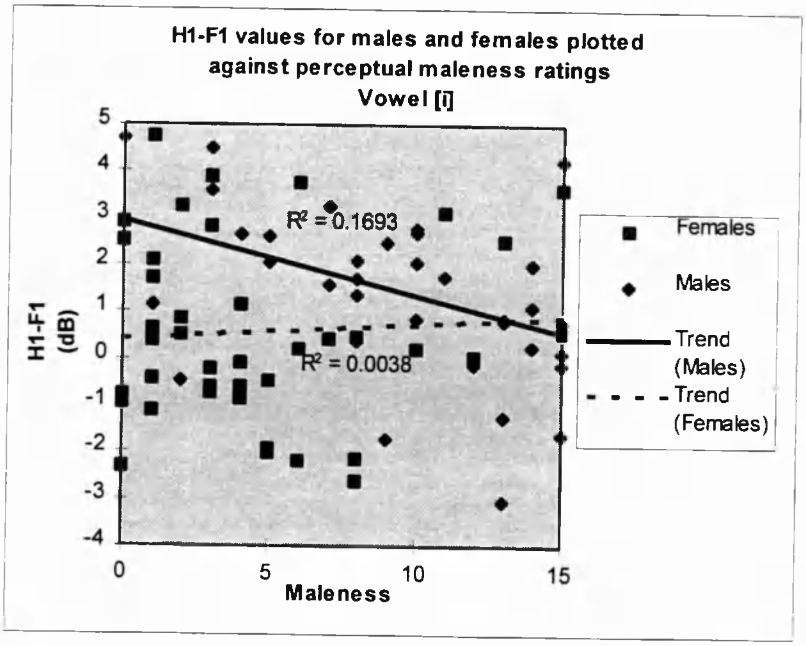
[o]	*least male-like*	*most male-like*
Mean	-10.445	-5.757
Variance	25.805	17.096
Observations	13	6
Hypothesized Mean Difference	0	
df	12	
t Stat	-2.132	
P(T<=t) one-tail	0.027	
t Critical one-tail	1.782	
P(T<=t) two-tail	0.054	**0.06>p>0**
t Critical two-tail	2.179	

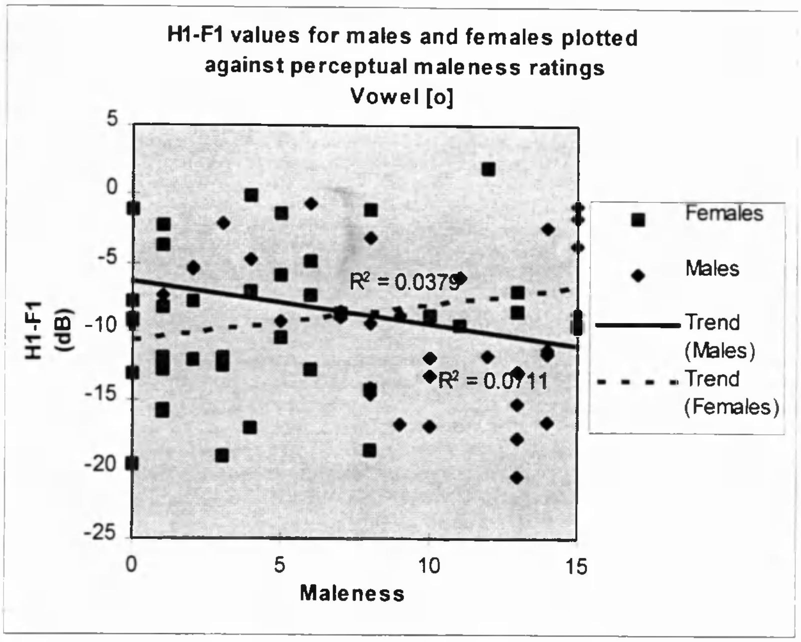
In order to observe any trend in the H1-F1 data with respect to the perceptual judgements made of the children, the values of the breathiness parameter of each child were plotted against the maleness rating. These plots are shown in figures 3.17 a-c.

Figure 3.17a, 3.17b and 3.17c (contd. overleaf)

H1-F1 for boys and girls plotted against number of judges responding 'boy' (i.e. maleness score). Also shown are R^2 values which indicate the regression coefficient of the best fit line through the data points.





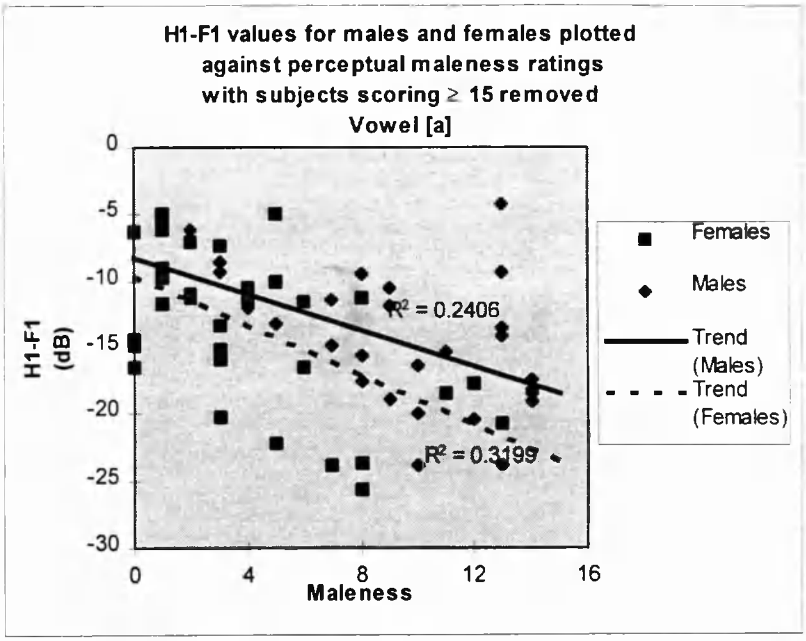


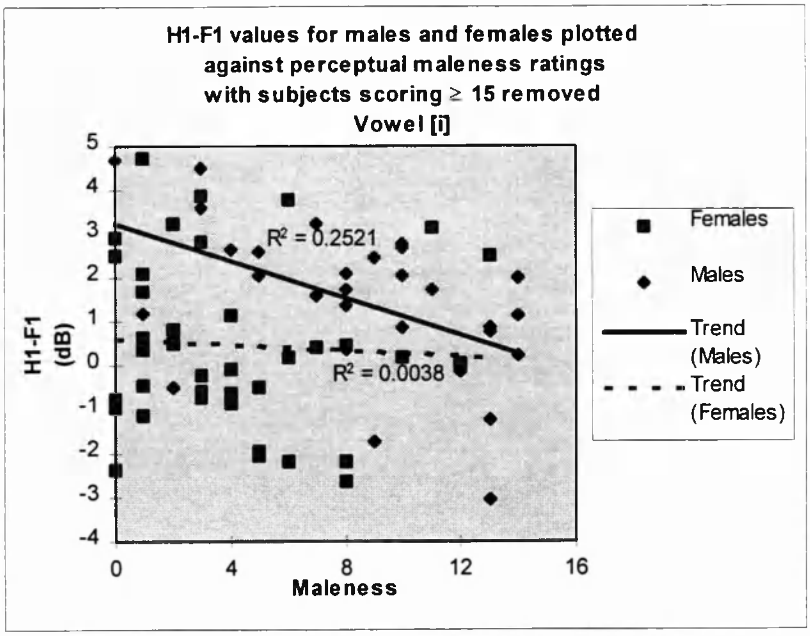
The results of the plot of breathiness vs. perceptual 'maleness' for /a/ indicates that breathiness is inversely proportional to perceived maleness. Vowel /a/ is of primary interest with respect to breathiness measures as explained in section 4.2.1.2 although the other vowels show similar trends for the male children at least.

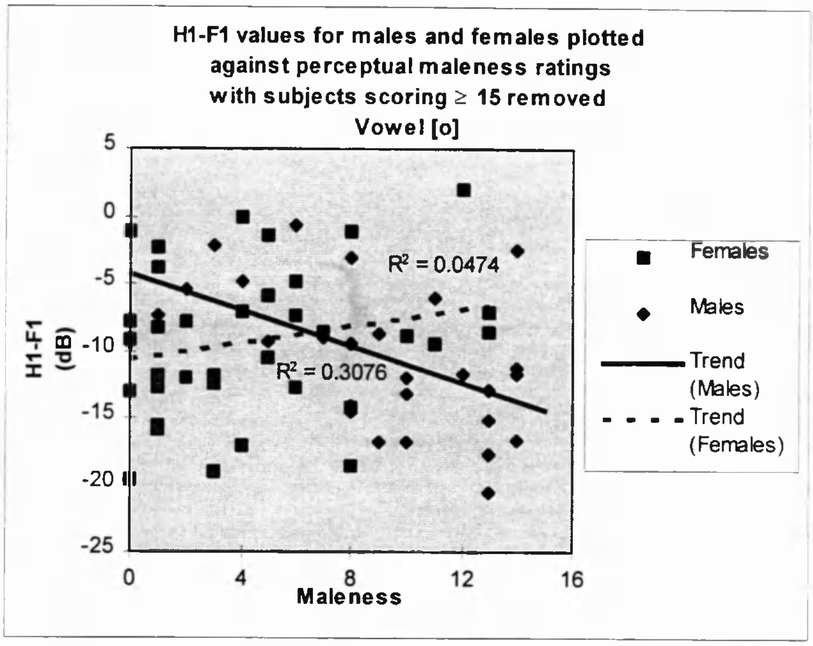
The data at the extreme positive end of the maleness scale, that is those children who were judged to be most male-like, show a selection of H1-F1 values which are spread over the whole range of values in the data pool rather than being confined to those values which the regression line would predict. This might suggest the operation of different cues used for identifying children who are more and less male-like. The charts have therefore been replotted with the data of the children who were judged to be most male-like (maleness ≥ 15) removed. These charts are shown in figures 3.18a-c.

Figure 3.18a, 3.18b and 3.18c (contd. overleaf)

H1-F1 for boys and girls plotted against number of judges responding 'boy' (i.e. maleness score). Also shown are R^2 values which indicate the regression coefficient of the best fit line through the data points. Subjects scoring ≥ 15 on the maleness scale have been removed.


H1-F1 values for males and females plotted
against perceptual maleness ratings
with subjects scoring ≥ 15 removed
Vowel [a]
H1-F1 (dB)
Maleness
R^2 = 0.2406
R^2 = 0.3199
Females
Males
Trend (Males)
Trend (Females)


H1-F1 values for males and females plotted
against perceptual maleness ratings
with subjects scoring ≥ 15 removed
Vowel [i]
H1-F1 (dB)
Maleness
R^2 = 0.2521
R^2 = 0.0038
Females
Males
Trend (Males)
Trend (Females)



3.2.2 Sentences

Table 3.29 shows the results of the acoustic analysis of the sentences of all 89 children. Means and standard deviations are displayed, as are the results of a t-test between the mean values of each parameter for males and females. The tests indicate that none of the acoustic parameters extracted by the MDVP differed significantly between boys and girls in this analysis (p>0.05).

Table 3.30 shows the result of the same set of acoustic analyses carried out over those children of either sex who were most often judged to be boys (*most male-like*) or girls (*least male-like*)[18].

The values of t displayed in Table 3.30 all fall below the critical value of t for p<0.05, indicating that there were no significant differences between these two groups for the parameters shown at this significance level.

[18] The children who form the categories 'most male-like' and 'least male-like' are listed in Appendix 8

Table 3.29 Acoustic measurements of the productions of the two test sentences showing the mean and s.d. for each parameter. The measurements are taken from actual boys and girls.

Acoustic parameter			GIRLS (n=43) Mean	GIRLS S.D.	BOYS (n=46) Mean	BOYS S.D.	t
Average Fundamental Frequency	**(Hz)**	***Fo***	263.15	27.59	263.78	24.25	-0.1584
Average Pitch Period	**(msec)**	***To***	3.88	0.37	3.86	0.36	0.3849
Highest Fundamental Frequency	**(Hz)**	***Fhi***	349.17	58.64	348.38	36.80	0.1049
Lowest Fundamental Frequency	**(Hz)**	***Flo***	191.12	24.21	190.01	28.51	0.2757
Standard Deviation of Fo	**(Hz)**	***STD***	26.75	12.19	25.29	8.86	0.8919
Phonatory Fo-Range in semi-tones		***PFR***	11.43	2.81	11.63	2.91	-0.4557
Length of Analyzed Sample	**(sec)**	***Tsam***	2.26	0.58	2.30	0.64	
Absolute Jitter	**(µsec)**	***Jita***	157.46	44.87	153.38	39.89	0.6281
Jitter Percent		***Jitt***	4.04%	1.07%	3.97%	0.95%	0.4585
Relative Average Perturbation		***RAP***	2.30%	0.60%	2.27%	0.56%	0.3213
Pitch Perturbation Quotient		***PPQ***	2.53%	0.71%	2.48%	0.60%	0.4710
Smoothed Pitch Perturb. Quotient		***sPPQ***	4.19%	1.57%	4.17%	1.36%	0.1316
Fundamental Frequency Variation		***vFo***	10.0%	3.44%	9.58%	3.24%	0.8250
Shimmer	**(dB)**	***ShdB***	1.25	0.36	1.20	0.31	1.0253
Shimmer Percent		***Shim***	11.4%	3.60%	10.8%	3.02%	1.1608
Amplitude Perturbation Quotient		***APQ***	10.9%	3.19%	10.7%	2.52%	0.5857
Smoothed Ampl. Perturb. Quotient		***sAPQ***	27.5%	6.67%	28.6%	4.87%	-1.1641
Peak-Amplitude Variation		***vAm***	51.0%	9.99%	51.5%	9.66%	-0.3121

Table 3.30 Acoustic measurements of the productions of the two test sentences showing the mean and s.d. for each parameter. Also shown is the value of *t* which resulted from a t-test between the acoustic values of most male-like and least male-like children's productions.

Acoustic parameter			LEAST MALE-LIKE (n=15) Mean	S.D.	MOST MALE-LIKE (n=10) Mean	S.D.	t
Average Fundamental Frequency	**(Hz)**	***Fo***	264.277	24.416	265.428	25.593	0.1179
Average Pitch Period	**(msec)**	***To***	3.85818	0.3009	3.84605	0.3857	-0.0870
Highest Fundamental Frequency	**(Hz)**	***Fhi***	358.444	65.642	356.397	33.279	-0.1133
Lowest Fundamental Frequency	**(Hz)**	***Flo***	192.905	20.030	189.173	26.182	-0.3964
Standard Deviation of Fo	**(Hz)**	***STD***	27.2766	15.332	26.8264	8.2459	-0.1045
Phonatory Fo-Range in semi-tones		***PFR***	11.75	2.9624	12	1.9771	0.2744
Length of Analyzed Sample	**(sec)**	***Tsam***	2.326	0.5944	2.04718	0.2269	*
Absolute Jitter	**(μsec)**	***Jita***	162.55	25.769	173.568	36.443	0.8551
Jitter Percent		***Jitt***	4.25%	0.85%	4.49%	0.71%	0.8383
Relative Average Perturbation		***RAP***	2.42%	0.49%	2.56%	0.39%	0.7975
Pitch Perturbation Quotient		***PPQ***	2.64%	0.58%	2.81%	0.46%	-0.8761
Smoothed Pitch Perturb. Quotient		***sPPQ***	4.31%	1.66%	4.57%	0.83%	0.5579
Fundamental Frequency Variation		***vFo***	10.09%	4.24%	9.90%	2.27%	-0.1592
Shimmer	**(dB)**	***ShdB***	1.32365	0.3002	1.42427	0.3281	0.8142
Shimmer Percent		***Shim***	12.09%	3.09%	12.63%	3.17%	0.4439
Amplitude Perturbation Quotient		***APQ***	11.53%	2.40%	12.06%	2.83%	0.5056
Smoothed Ampl. Perturb. Quotient		***sAPQ***	28.59%	3.33%	30.01%	5.23%	0.7830
Peak-Amplitude Variation		***vAm***	52.67%	7.27%	53.68%	7.35%	0.3559

Chapter Four
Discussion and Further Analysis

4.0 INTRODUCTION

Following the structure of the previous chapters, the discussions of the perceptual and acoustic phases of the experiment will be treated separately. Section 4.1 presents a discussion of the results of the perceptual study including comments on the rates of identification encountered and the measures of bias. Section 4.2 discusses the findings of the acoustic study and includes commentary on the comparisons of acoustic parameters. Finally, there is a general conclusion and summary in section 4.3 which draws together the overall findings of the experiment and locates them with reference to existing theories of child speech development.

4.1 PERCEPTUAL STUDY

4.1.1 Differences in rates of identification between samples

The rates of correct identification of gender in the prepubertal children of this study compare favourably with rates reported throughout the relevant literature. Tables 4.1-4.4 present a visual comparison of the present results with some previous findings.

There are very few studies in which English speaking children as a group have not been significantly well identified (e.g. Sachs, 1975 [backward sentences only]). The main finding that all of the samples were recognised at rates significantly above chance is therefore not unexpected. The anticipated ability to identify the gender of children prior to puberty from samples of voice alone was borne out.

Table 4.1 A comparison of the results of gender identification studies which have used isolated vowels as samples.

Vowel Samples Used

Investigator(s)	Age of Children	Correct Identification Rate
THIS STUDY (1995)	4-5 years	66%
Bennett and Weinberg (1979a)	6-7 years	66%
Sachs (1975) / Sachs et al.(1973)	4-14 years	66%
Gunzburger, Bresser & Ter Keurs (1987)	7-8 years	60%
Bresser & Gunzburger (1985)	7-8 years	55%

Table 4.2 A comparison of the results of gender identification studies which have used sentences as samples.

Sentence Samples Used

Investigator(s)	Age of Children	Correct Identification Rate
Sachs et al.(1973)	4-14 years	81%
THIS STUDY (1995)	4-5 years	76%
Bresser & Gunzburger (1985)	7-8 years	74%
Bennett and Weinberg (1979a)	6-7 years	71%
Ingrisano, Weismer & Schuckers (1980)	4-5 years	71%
Sachs (1975) / Sachs et al.(1973)	4-14 years	59% (backward sentences)

Table 4.3 A comparison of the results of gender identification studies which have used spontaneous speech as samples.

Spontaneous Speech Sample Used

Investigator(s)	Age of Children	Correct Identification Rate
Weinberg and Bennett (1971)	5-6 years	74%
THIS STUDY (1995)	4-5 years	73%
Karlsson and Rothenberg (1992)	3-8 years	69%

Table 4.4 A comparison of the results of gender identification studies which report overall results only

Other Studies (overall results)

Investigator(s)	Age of Children	Correct Identification Rate
Ingrisano & Thompson (1975)	4-5 years	70%
Marshall (1972)	8-9 years	69%

As remarked upon in chapter one, the differences in recognition rates evidenced across the different studies within each category are most likely attributable to differences in methodology combined with age differences in the subjects used. There does not appear to be a clear age-related trend in the rates of correct identification, however, as there was a range of different methodologies and as the pools of children were often not balanced between boys and girls, we cannot use the apparent ranking of the studies as an indication of the effect of increasing age on perceptual identification of gender from voice. A more consistent experimental paradigm applied across children of different, non-overlapping age groups is required to illuminate the effect of child age on gender identification.

What is of more interest are the differences between the samples. In this study, the sentence and (probably) the spontaneous speech samples were both identified at higher rates than the vowel sample[19]. Judging by the patterns of previous results seen above, this is to be expected. At least two possible explanations for this finding are suggested. The first explanation is that sentences and spontaneous passages clearly contain more phonetic information than isolated vowels. Whatever the acoustic / phonetic markers of gender that are present in the isolated vowels, they must also be present in the longer sample types and in greater quantities. This difference in **quantity** of gender cues may be sufficient to account for the differences discovered between the sample types. Alternatively, or additionally, there may be different cues which are specific to connected speech samples present in the sentences and passages. This difference in **quality** of gender cues may be the determining factor in the rates of identification.

As the isolated vowel sample was still recognised at above chance levels of accuracy, it would appear that there must be at least some spectral information which varies relatively consistently according to gender.

[19] The Scheffé tests indicated that vowels are dependably worse identified than sentences and almost, but not quite, dependably worse identified than passages. There was no significant difference between the rates of identification of the sentences and passage samples.

The most likely explanation for the identification rate differences then is that there are one or more basic acoustic cues which listeners can use to identify gender of a speaker which are present in even single, short stretches of phonation (i.e. vowels), but there may be other features which are encoded into longer sections of speech which enhance the listener's ability.

A second interesting point about the sample identification rates is the fact that there was no significant difference between the correct judgements using the sentence and spontaneous speech samples. The result from Sachs *et al.*(1973) which reported 81% correct identification on the basis of sentences must be questioned as some of their child subjects were as old as 14 years and had probably already entered the first stages of puberty. The anatomical sexual dimorphism associated with puberty would already have started to separate the genders quite markedly at this stage and the vocal organs of girls and, particularly, boys would have begun to enlarge rapidly. A consideration of the rates reported in previous studies (tables 4.1 - 4.4) reveals an approximately similar level of listener accuracy between sentence and spontaneous passage samples. The question is raised "why does identification of gender not improve as we move from sentence to spontaneous speech samples?". After all, both sample types contain approximately equal amounts of phonetic information, both are examples of running speech and the spontaneous speech sample is the most life-like in that it resembles actual communicative output the most closely. Furthermore, the spontaneous sample allows the child to express him or herself freely by using whatever syntactic, morphological, semantic, pragmatic and phonetic constructions he or she pleases. It is unconstrained (except for the broad subject matter) and hence represents a free-choice situation in which we would expect many of the child's sociolinguistic vocal markers (such as age, health, socio-economic status) to rise to the surface. Gender is a prime exemplar of such social indices and therefore would be expected to be revealed to the fullest degree when the speech patterns aligned the most closely with normal modes of everyday communication.

The expectation, then, is that the spontaneous speech passages ought to offer the listener the best opportunity of identifying a speaker's gender, followed by the sentence sample, followed finally by the vowel sample. It is slightly surprising therefore that in this study, as in other, previous studies, the spontaneous speech was not the sample from which the listeners gained the most gender-bearing information.

An explanation for this situation in the present study can be reached by considering the procedure involved in collecting each of the samples. As a reminder, the spontaneous speech was collected by asking the child to recount a story which he or she had been told immediately previously using pictures as a guide. The topic of speech was therefore constrained although the actual phonetic content of each child's utterances (i.e. the words used by the child) was open to individual variation. In the sentence condition, the child repeated two sentences which he or she had learned to associate with pictures. Each child spoke the same two sentences. From the viewpoint of the listeners therefore, there was an additional stage of mental processing involved in the gender identification process for the passage sample - that of normalising or extrapolating the child's gender from continually varying phonetic input. In other words, when the listener heard a child speaking a spontaneous passage they would first have to neutralise the effect of the different constructions used before they could extract the gender-bearing information. In the sentence sample however, the listener could directly compare like-with-like and the only apparent variation between the children's productions was the influence of the child's voice.

The fact that the listener did not use the potential differences in syntactic or lexical constructions, which boys and girls may have employed, as a cue for gender recognition is suggested by the fact that the level of accuracy of identification did not increase from sentence to spontaneous speech samples.

It is hypothesised, then, that the extra cognitive processing stage which a listener must apply to varying spontaneous speech data in order to extract comparative gender information serves to reduce the levels of accuracy of gender identification by placing a greater demand on the perceptual systems of the listener.

4.1.2 Gender differences in identification rates

The issue of gender differences between the perceptual abilities of the adult listeners and also between the perceptibility of the children has not been treated systematically across previous research. Many studies have only included either male or female listeners in their judging panel and, of those which have included both sexes, most have failed to balance their numbers.

Table 4.5 Correct recognition rates of boys and girls in those experiments in which results are separated by gender.

Investigator(s)	Children		Best identified sex
	Boys	Girls	
VOWELS			
Sachs (1975) / Sachs et al.(1973)	72%	59%	M > F
Bennett and Weinberg (1979a)	68%	63%	M > F
THIS STUDY (1995)	63%	69%	F > M
Gunzburger, Bresser & Ter Keurs (1987)	57%	52%	M > F
SENTENCES			
THIS STUDY (1995)	73%	79%	F > M
Gunzburger, Bresser & Ter Keurs (1987)	71%	76%	F > M
Ingrisano, Weismer & Schuckers (1980)	*	*	F > M
Bennett and Weinberg (1979a)	71%	69%	M > F
Sachs (1975) (backward sentences)	69%	49%	M > F
SPONTANEOUS SPEECH			
THIS STUDY (1995)	70%	75%	F > M
Karlsson and Rothenberg (1992)	68%	71%	F > M

It is clear from Table 4.5 that there is not a universal pattern to the gender differences in accuracy of child identification by voice. Some studies show boys to be better identified whilst others show that the girls are best identified. This study found that girls were identified more accurately than boys in all samples. Furthermore, this study provides the only report of higher identification rates of females using isolated vowels.

Two analyses of variance were used to determine whether there was a statistically significant difference between the identification rates of boys and girls (one using all judges and one with judge f1 removed).

As judge f1 was discovered to be an outlier, the data and analyses with f1 removed from the calculations are considered to be better indicators of the general performance of the populations from which our experimental samples were drawn.

When we consider the sex of the child as an effect, i.e. the degree to which the sex of the child influenced the judges' responses, it is clear that there is a real difference in the performance of judges between boys and girls. An examination of the cell means or the table of percentage correctly identified children reveals that girls were correctly recognised more often than boys. We now know that this difference is a statistically significant one. It is interesting to notice also that the removal of only one judge from the original analysis alters the outcome of the ANOVA from non-significant to significant.

Table 4.6 p-values of the F-Ratios of two sets of ANOVAs performed on correct gender identification rates. Shaded cells indicate significant main effects ($p<0.05$).

	All Judges		Without Judge f1	
	By children	By judges	By children	By judges
Sex of Child	0.087	0.097	0.012	0.0001
Sex of Judge	0.014	0.297	0.002	0.161
Material type	0.027	0.0001	0.027	0.0002

The effect of the sex of the judge is less straightforward. We can see the same pattern of results in table 4.6 irrespective of whether we remove judge f1. The results appear to indicate that there is a significant difference between male and female judges in the 'By Children' analyses but not in the 'By Judges' analyses. Which result do we believe and why are they different?

The answer to the first question is that we believe both results and the answer to the second is 'because they are telling us different things'. The difference between the two results is the reason that we perform two ANOVAs in the first place. To put it simply, the variance is different in the two cases. In the 'By Judges' case, the individual judges were so variable that we cannot say that the difference between male and female judges overshadowed the variability in each group. In the 'By Children' case, the situation is different. While individual judges varied to a reasonably large degree, individual children's effects on male and female judges were much more stable. In other words, the variability in correct identification scores among the children was small relative to the judge sex effect.

It is clear therefore that the significance or lack of significance in the sex of judge variable depends on the ratio of the variances of the correct identification scores with respect to both judges and children and as this ratio differs according to whether we focus our attention on judges or children, the outcome is two separate (and apparently conflicting) results. However, given the fact that the variability amongst the judges is so great that it swamps any effect of sex in the 'By Judges' analyses, we must assume that there is no statistical difference between the performance of the male and female judges.

Although the removal of judge f1 from the ANOVA did not alter the outcome of the analyses in terms of significance, the p-values of the 'sex of judge' variable were lower when f1 was removed from the statistic. This seems to indicate that the male and female judges were more distinct when the outlier was removed although the effect of losing f1 was not marked enough to cause an important shift in the results. This finding is to be expected. We already know that the female judges tend to perform slightly more accurately than the male judges and table 3.9 in section 3.1.2.2 indicates that following the removal of f1's data, the correct identification rates of the female judges increase. As the female judges already have a slight advantage over the males, this further improvement in the identification rates represents a widening of the performance gap between the sexes of judges. In other words, the levels of ability shown by male and female judges became more distinct following the removal of f1.

The issue of why females are better identified than males in this study is difficult to answer. The first task is to determine whether this finding is a real result or merely an artefact of the experimental design or statistical analysis.

The data which forms the basis for the identification rates is the actual judges' responses. If there were no listener bias operating in the judgement process, we would expect the responses to be split equally between 'boy' and 'girl' judgements. As there were more boys than girls participating in the study (51.7% vs. 48.3%) we would anticipate that the responses should approximately match these percentages. The expected response split and the actual observed response split are shown below.

Table 4.7 A comparison of the observed and expected number of 'boy' and 'girl' responses for each sample type.

	'BOY'	'GIRL'
Percentage of all subjects	51.7%	48.3%
VOWELS (N=1335)		
Number of expected responses	690.195 (51.7%)	644.805 (48.3%)
Number of observed responses	619 (46.4%)	716 (53.6%)
SENTENCES (N=1335)		
Number of expected responses	690.195 (51.7%)	644.805 (48.3%)
Number of observed responses	615 (46.1%)	720 (53.9%)
SPONTANEOUS SPEECH (N=1335)		
Number of expected responses	690.195 (51.7%)	644.805 (48.3%)
Number of observed responses	625 (46.8%)	710 (53.2%)

As described in section 3.1.2 there is a detectable bias amongst individual listeners in the judgement results. Listeners (with the exception of judge f1 who was removed from the analyses) were significantly biased to respond 'girl'. This is reflected in the overall percentage of 'girl' responses (53.6%). This apparent bias is further increased by the unequal proportion of boys and girls participating in the study.

The isolation of the bias in the response judgements does not, however, in itself answer the question regarding the nature of the difference in the identification rates between boys and girls. It has been established that there are more 'girl' responses than expected by chance and that there is a bias operating. The original question might now be more usefully expressed as 'what underlies the perceptual bias?'. Is it that listeners are basing their 'girl' responses on actual vocal characteristics of the speakers or are they based on some imaginary preconceptions of the judges?

If part of the observed bias were based on the 'real' ability to identify gender then we would expect there to be some relationship between the 'girl' response and one or more acoustic parameters. The comparison of most male-like and least male-like children with respect to the acoustic parameters extracted was carried out and is reported in table 3.30 of section 3.2.2. There are no significant differences between the values of any of the parameters for the two groups of children. In other words, at first sight, there does not appear to be a relationship between the number of 'boy' or 'girl' responses (i.e. the group to which the child belongs) and any acoustic parameter. However, the reader should remember that a failed attempted correlation does not necessarily mean that there is no correlation in reality. If the speakers displayed a small range of values of the parameter in question then a statistical correlation might fail although the two samples may be correlated in reality.

Appendix 9 shows a sample of cumulative frequency histograms which show the distribution of values of selected acoustic parameters for boys and girls. They clearly show that the data is normally distributed and that, generally, a large proportion of the values fall in a relatively small central range. For example, considering the data for mean fundamental frequency, 64% of the boys' Fo values and 72% of the girls' values fall within the range 225-275 Hz and almost 90% of the values for both sexes fall within the range 225-300 Hz. When we look at absolute jitter we find that 77% of the boys' values and 72% of the girls' values fall within the range 100-200 μs and 92% of the boys' values and 89% of the girls' values fall within the range 100-250 μs. Finally, considering percentage shimmer we see that over 81% of the boys' values and in excess of 90% of the girls' values fall within the range 0.75-1.75%. The finding that the acoustic data is relatively tightly clustered lends support to the belief that the lack of correlations in the results is partly due to the narrow range of output values displayed.

A likely explanation for the pattern of perceptual results, given the available data, is that listeners were making use of both of the discrimination strategies open to them.

They were using one or more acoustic / phonetic cues as a guide to the gender of the speaker but due to the small range of values used by the children, it was impossible to isolate these and, at any rate, these cues were not always sufficient to uniquely identify a child's gender. Listeners are likely to have also brought with them into the experiment a preconception about children's voices. It seems likely that the reason that considerably more 'girl' responses were recorded than 'boy' responses is that the response of "female" or (in this case) 'girl' was the default value. All prepubertal children have voices which sound more like that of an adult female than an adult male and listeners may have been predisposed to respond 'girl' unless there was overwhelming evidence to the contrary. In effect, the bias which is apparent in the perceptual responses may represent a categorisation of those 'in-between' children (i.e. those children for whom the listeners were unsure) as girls. In such a case, the responses would pattern as follows: children who were recognised as females = 'girl'; children who were recognised as males = 'boy'; children about whom the judges were unsure = 'girl' (default value). This would lead to good gender identification rates in general and more 'girl' responses than 'boy' responses. This is the situation evident in the results of this experiment. It would also account for the higher rate of identification for girls. A higher proportion of 'girl' responses across all of the children will result in a higher correct identification rate for girls.

The study carried out by Ingrisano, Weismer and Schuckers (1980) is perhaps the only one of its kind which shows a sex-difference in ability to judge the gender of prepubertal children. As a reminder, their results show (1) female listeners were more accurate than male listeners in identifying male children's voices; (2) male listeners were more accurate than female listeners in identifying female children's voices; (3) female voices were more accurately identified than male voices by both listener sexes and (4) female listeners were better judges of child sex than male listeners, when data were collapsed across child sex. Let us compare these results with the findings of the present study[20]. The first result of Ingrisano *et al.* (1980), that female listeners were more accurate than male listeners in identifying boys' voices, is consistent with our findings as presented in table 3.9, section 3.1.2.2 (see also fig. 3.9). In our case, the difference in number of correct boy judgements between female and male listeners is statistically significant (p=0.01).

The same data also accounts for the fourth finding of Ingrisano *et al.* (1980). When the difference in accuracy of our male and female judges is assessed across all children by collapsing child sex, females appear to be slightly (although not significantly) more accurate (p=0.161).

The third finding of Ingrisano *et al.* (1980), that female voices were more accurately identified than male voices by both listener sexes, has already been discussed above and is also consistent with the present findings - this leaves their second finding. The second finding states that male listeners were more accurate than female listeners in identifying female children's voices. This is in direct contrast to the results of this experiment which found that female listeners were better than male listeners at identifying gender <u>irrespective of the sex of the child</u>.

4.1.3 Bias and Discrimination

As described in section 2.1.5.1, measures of statistical bias and discrimination will tell us something about the ability of our listeners to correctly identify the voices of the prepubertal children. The measure of discrimination (d') gives us a direct indication of how accurate a listener is and allows us to compare individual listeners and groups of listeners (e.g. male judges against female judges). The measure of bias (log β) shows us to what extent a listener or group of listeners is predisposed to respond in a particular manner. In this experiment we can see if there is a bias towards responding either 'boy' or 'girl' on the part of the listeners and also if there are any listeners who are reducing the validity of the group mean data due to their extreme values of log β (i.e. outliers).

[20] The results referred to in this section are those analyses which have had the data from judge f1 removed. The rationale behind this step is fully explained in section 4.2.

4.1.3.1 Bias

The notion of bias in perceptual experimentation is an important one. In attempting to quantify and statistically analyse individual subjects' responses we need to be sure that the apparent variations in the data with which we are dealing are indeed the variations which are of interest. As an analogy, consider that favourite example of statisticians, the coin toss. There are two possible outcomes of the coin tossing process - 'heads' and 'tails'. If we could toss a coin an infinite number of times we would expect exactly half of the results to be 'heads' and exactly half to be 'tails'. Unfortunately, the situation where there are an infinite number of coin tosses (or subjects) never occurs. Consider therefore the situation where we toss a coin 1,000 times. According to probability theory we would expect the outcome to be 500 'heads' and 500 'tails'. If this did occur we would be satisfied that there was no bias in operation. What if the outcome was 501 'heads' and 499 'tails'? We would probably still believe the situation to be unbiased. If the outcome was 700 'heads' and 300 'tails' we would quite correctly believe that there was probably[21] a bias towards the outcome 'heads'. How far do we have to deviate from the optimal position of equal response ratios for each equally probable outcome before we recognise bias?

This question is answered by the theory of signal detection. In terms of this experiment, we want to know whether our male and female judges are biased in their responses, if so, whether they show a preference for responding "boy" or "girl" and whether any bias which exists is present over all sample types.

We can see from figure 3.5 in section 3.1.2.1 that in all cases (except the female judges' responses in the sentence sample condition which, as we will shortly see, is a special case) the judges have mean values of log β in excess of 0. This indicates that they are biased to respond "girl". Both male and female judges have positive values of log β and both groups are therefore biased. Male judges have greater mean values of the statistic which indicates that they were even more inclined to respond "girl"

[21] There is always a chance, however infinitesimal, that an outcome such as 700 'heads' and 300 'tails' will occur in an unbiased system, however, given an unlikely result, the chance that there is a bias operating increases as the chance of that particular outcome decreases.

than the female judges. As is stated in section 3.1.2.1, the values of log β of the male judges were significantly greater than 0 whilst the values for the female judges did not differ significantly from chance, however it is clear that the pattern for all judges is to show a tendency to respond “girl” rather than “boy”. This result is the statistical reason behind the different numbers of “boy” and “girl” responses described in section 4.1.2.

The pattern in log β values visible in figure 3.5 is dramatically spoiled by the level of the female judges in the sentence sample condition. Reading this data, we would be lead to think that the female judges are suddenly and unexpectedly reversing the overall bias of answering in favour of girls and actually favouring boys. This fact, in addition to the suspiciously large standard deviation value for this condition, leads us to inspect the data more closely. When we examine figure 3.3b which shows the individual values of log β for the female judges in the sentence condition, we see that most of the judges, with one notable exception, have log β values which are clustered from 0 to 0.25. In other words, most of the female judges do in fact display a small bias in favour of a “girl” response. The exception is judge f1 who has a value of -0.91. This value represents a massive bias in favour of a “boy” response and is clearly unrepresentative of the remainder of the female (or male) judges. The value of -0.91 is so far from the mean value of the other judges that it can be considered an outlier and is a candidate for deletion. Before disregarding f1’s data however, let us examine her values of log β for the other samples. It so happens that for the vowel sample she has log β = -0.13 and for the passage sample the value is -0.11. It is clear therefore that this individual judge is biased throughout the whole experiment in favour of responding “boy” and, as such, she is seriously affecting the mean values of the bias data. When we remove the data from f1 we see the pattern of log β values which we would predict (figure 3.10, section 3.1.2.2) - the values of log β are more similar for male and female judges, there are no negative mean values of log β and, on this occasion, log β of both male and female judges are significantly greater than 0 (with the exception of male judges in the sentence sample which is due to the rather large amount of variance in this data).

The values of p in the two ANOVA result tables reinforce the belief that the deletion of judge f1 from the analyses yields a more realistic result. In table 3.6 of section 3.1.2.1 which shows the ANOVA carried out on the values of log β for all judges (including f1) the p-value for the main effect of sample type is 0.875. In table A6.1 of appendix 6 which shows the result of a similar ANOVA without the data of judge f1, the p-value for the main effect of sample type is 0.129. This indicates that with the outlier data of judge f1 in place the judges' mean bias values across all sample types were very similar. After the outlier data had been removed the similarity of the bias scores between samples dropped considerably ($p=0.129$) although the difference was still not quite statistically significant. A large part in this modification must be played by the reduction in variance of the log β values of the female judges' sentence sample. The large variance caused by the data of f1 (visible in figure 3.7, section 3.1.2.1) makes it less likely that an *apparent* difference between the mean values of log β across the different sample types will reflect an *actual* statistical difference. With the data from f1 removed (and hence a large proportion of the variability removed) the differences in the mean values of log β are considerably more likely to represent a real difference between the samples (although, as the result stands with $p=0.129$, the probability that the means are drawn from samples of the same population is about 13 in 100 which is not low enough to reject the null hypothesis). The shift in probability of a difference between the log β values of male and female judges after the removal of f1's data is also interesting. Before removal the p-value is 0.317 and after removal the figure rises to 0.964, indicating that, as expected, the log β values of male and female judges become more similar after the removal of the outlier.

4.1.3.2 Discrimination

The statistical measure of discrimination used in this study is d' and, as stated earlier, we can think of d' as a measure of a judge's ability to discriminate noise from signal (i.e. girl from boy). Mathematically d' is characterised in terms of the distributions of the signal and noise values of the perceptual response - specifically, d' is the distance of the signal mean from the noise mean in multiples of the noise standard deviation. Therefore d' can be thought of as a standardised score (or z-score).

Considering figure 3.6, section 3.1.2.1, we can immediately recognise that the pattern of d' values matches the rates of correct identification shown in table 3.1, section 3.1.1. In both cases female judges score more highly than male judges and in both cases the sentence sample attracts the highest scores, followed by the passage sample and the vowel sample scores least highly. From first appearances, therefore, it seems that d' is fulfilling its promise as a reliable measure of accuracy of gender identification.

The differences in identification rates between the sample types were measured by t-tests and reported in table 3.3 of section 3.1.1. and were also measured by ANOVA and reported in table 3.4 of section 3.1.1.1 and table 3.5 of section 3.1.1.2. They indicate that the sentence sample was identified significantly more accurately than the vowel sample and that the difference between the passage sample and vowel sample at $p=0.08$ marginally failed to achieve significance. There was no significant difference between the rates of identification of the sentence and passage samples.

Sign tests were used to determine the differences between the d' values of the judges across the different samples (tables 3.8a-c, section 3.1.2.1). These non-parametric statistics confirm the findings of the parametric t-tests. The sign test between judges' d' in vowel and sentence samples showed a clearly significant difference ($p<0.05$) whilst the same test between d' values of sentence and passage samples showed no significant difference. The third test between d' values of the vowel and passage samples produced a marginal failure. The test statistic in this case was exactly equal to the critical value of r indicating that the result approaches significance at the $p=0.05$ level. In other words, the probability of the d' values of these two samples being from the same population is almost exactly 5% (0.05) which is not enough to reject the null hypothesis outright, but enough to indicate a very strong likelihood that the figures represent different results. This, of course, matches the result of the t-test which indicated that the probability of the recognition rates of the two samples being the same was 8% (0.08).

The removal of the d' data of the biased judge f1 did not change the overall pattern of d' results. The sentence sample remains the best identified and the vowel sample remains the worst identified. The females are still better judges than the males. The ANOVA summary in table A6.2 in appendix 6 shows that after removing the outlier the d' values of the different samples are even more significantly different and that the d' values of the male and female judges are slightly more similar. This indicates that the d' data from judge f1 was 'artificially' drawing the mean d' levels of the different samples together. The remainder of the 'unbiased' d' data is distributed in such a way as to separate the samples, and even with judge f1 included they are diverse enough to cause a significant difference between the mean d' of the different samples, however with judge f1 removed this difference is increased.

The fact that the removal of judge f1's d' data did not cause as large a shift in overall pattern as was the case with the log β data is due to the fact that although strongly biased, judge f1 displayed relatively normal values of d'. This would seem to indicate that although strongly inclined to respond 'boy' in cases where there was a doubt over the gender of the subject, f1 was still achieving a good rate of correct identification. These facts are not, in fact, contradictory. Whereas other judges might respond correctly to some boys and correctly to some girls and then more or less balance their incorrect responses between boy and girl answers (with a small bias towards 'girl'), f1 would respond correctly to both boys and girls and when making errors would be much more likely to respond 'boy'. This would lead to the situation where she has a good d' (due to the correct responses made) but a large negative log β (due to the relatively large proportion of 'boy' responses made).

4.2 ACOUSTIC STUDY

4.2.1 Vowels

4.2.1.1 Formant measures

The vowel formant results obtained from the child subjects in this study concur reasonably closely with other published data (Bennett, 1981; Eguchi and Hirsh, 1969) although a large amount of inter-subject variability has been reported in all studies which have attempted to measure children's formant frequencies. The results of the statistical analyses comparing formant measurements of the boys' and girls' vowel productions of the present experiment are summarised in table 4.8. It is clear from this data and from the raw formant frequency measures (tables 3.13-3.15, section 3.2.1.1) that the girls tended to have higher formant frequencies than the boys although this was not always statistically significant. Of the nine comparisons (F1, F2, F3 of three vowels), three were clearly significantly different, two others were marginal failures of significance and the remainder showed no significant difference. Viewed as a whole, eight of the nine comparisons reveal that the girls have higher mean formant frequencies than the boys. These eight comparisons also yield p-values less than 0.25 which leads one to believe that the formant frequency data is not randomly distributed between the sexes but is beginning to show the rise of the sexual dimorphism which is responsible for almost perfect gender recognition among adult speaker and listener pairs. There is still a degree of overlap in the formant values, and the relatively large variability displayed by these child speakers, particularly in F3, tends to mask any specific gender differences, however the trend of female speakers showing higher formant frequencies appears to have emerged by the age of these subjects (about 5 years).

Table 4.8 Results of statistical comparisons of the first three formant frequencies, bandwidths and amplitudes of boys' and girls' productions of 3 sample vowels. All tests were independent t-tests assuming unequal variances.

	a	**i**	**o**
Formant 1	p=0.015 *	p=0.120	p=0.145
Formant 2	p=0.007 *	p=0.016 *	p=0.055 (*)
Formant 3	p=0.206	p=0.077 (*)	p=0.843
Bandwidth 1	p=0.395	p=0.479	p=0.458
Bandwidth 2	p=0.540	p=0.796	p=0.148
Bandwidth 3	p=0.610	p=0.244	p=0.087 (*)
Amplitude 1	p=0.011 †	p=0.097 (†)	p=0.072 (†)
Amplitude 2	p=0.025 †	p=0.760	p=0.298
Amplitude 3	p=0.059 (†)	p=0.379	p=0.034 †

* = Values of girls formant measure > values of boys' formant measure
† = Values of boys' formant measure > values of girls formant measure
Parentheses indicate marginal failures of statistical significance in the direction indicated (0.1>p>0.05).

The absolute differences between the mean formant frequencies of the boys and girls in the present study were smaller than those reported by Bennett (1981) also, in general, the formant frequencies of the children used in this study are higher than those of Bennett's subjects. As the children used in Bennett's study were 7 - 8 years old, that is an average of 2½ years older than the children in this study, this finding can be interpreted as support for two aspects of the evolution of the adult sexual dimorphism. First, there is further evidence that a normal developmental growth pattern will result in a progressive lowering of formant frequencies as the vocal tract lengthens. There are many assumption that must be made if we are to compare the formant data of Bennett's American children with the present data, such as accent differences, possible differences in developmental rates, socio-economic and cultural differences in the child subjects. Leaving these potentially influential issues to one side, if we compare the mean formant frequencies of the boys and girls from the two studies, it would seem that in the space of 2½ years most of the values of /i/ and /a/ [22]drop substantially. The average size of the drop is 323 Hz for F2 of /i/, 125 Hz for F3 of /i/, 452 Hz for F1 of /a/ and 328 Hz for F3 of /a/. F1 of /i/ appears to rise very slightly and there is no clear difference in F2 of /a/. It is also difficult to compare the present findings with the findings of Sachs *et al.* (1973) as the latter used children whose ages ranged from 4 to 14 years and therefore the oldest of these children would be entering the pubertal development stage; however the average age of Sachs et al's subjects is clearly greater than the average age of this study's subjects (5 years). The mean F1 and F2 values of the vowels /i/ and /a/ were considerably higher for the children in the present study than for children in Sachs *et al's* study.

Second, there would seem to be an indication that the vocal tract growth of boys and girls prior to puberty is disproportionate. We know that adult males and adult females display differences in formant frequencies and that these differences are mostly due to the longer vocal tract in males. It is less clear, however, whether these differences manifest themselves primarily at the time of puberty when the male growth spurt is greater than the female's or if the differences are cumulative from

[22] Note that whilst the present study assigns the symbol /a/ to the vowel quality, Bennett uses the symbol /æ/.

birth. Many researchers report only single figures for average formant frequencies of children and do not separate the children by gender prior to puberty, claiming that there are no significant differences in this period (Eguchi and Hirsh, 1969; Peterson and Barney; 1952). The findings of Bennett (1981), Sachs *et al.* (1973) and this study each lend support to the notion that the sexual dimorphism evident in adult voices has its inception well before the onset of puberty. There appears to be a developmental trend in inter-gender formant differences amongst the three studies listed. The study which uses the youngest children (the present study) reports the smallest sex-differences in formant frequencies, Bennett's study, which uses children who are on average 2½ years older, reports larger sex-differences and Sachs *et al.*, who use a large age range report variable sex-differences amongst their formant results.

Fant (1966; 1973; 1975) has written more than anyone on the subject of formant scaling and the contribution of various sections of the vocal tract on age and sex-differences in formant frequencies. Using Fant's *k*-factors as an indication of sex-differences in formant frequencies (see section 1.1.5 and later in this section) the developmental trend becomes clear. The larger the value of *k*, the greater the sex-difference in formant frequency. When averaged across F1-3 of five vowels, the mean value of *k* (i.e. the sex-difference) was 19% for Peterson and Barney's (1952) adult English speakers and 18% for Fant's (1973) adult Swedish speakers. Bennett's (1981) 7-8 year old English speaking children yielded *k*-factors of around 10%. Sachs *et al.'s* (1973) subjects who varied widely in age demonstrated *k*-factors which ranged from 0.1% to over 12%. When averaged across the three vowels used in the present experiment, our children exhibited a mean *k*-factor of 3.77%.

The indication would seem to be that there are small but perceptible differences between the formant frequencies of boys and girls by the age of 5 years and that these sex-differences continue to grow in the years before puberty. It is not possible to make any inferences regarding vocal tract lengths of these children. Several studies have published anatomical measurements which indicate that there are negligible length differences in the supraglottal vocal tract of prepubertal boys and girls (Walker and Kowalski, 1972; Hunter and Garn, 1972). It may be, therefore, that the formant differences seen in this experiment and in those cited above stem largely from sex-specific phonetic manipulations of the articulators in order to increase the perceived maleness or femaleness of the speaker in the same way that adult speakers use such a technique to enhance the biological dimorphism (see Mattingly, 1966). The alternative is that there are indeed small differences in vocal tract length which continue to widen as the prepubertal child develops and that earlier studies failed to isolate the crucial measurements. At puberty itself, of course, there is a larger, unequal and quantal increase in vocal tract length in both boys and girls which results in the final separation of the male and female formants.

The role played by formant bandwidth and amplitude in the gender-identification process has been neglected in comparison to the amount of attention heaped upon formant frequencies. Also, because they are partially interdependent, they are often not quoted separately in reports of acoustic measurements.

There is far less of a distinction in the present formant bandwidth data than in the amplitude data. From table 4.8 we see that none of the nine comparisons between mean formant bandwidths of boys and girls were significantly different. Upon inspection of the raw results we discover that the girls tend to have larger bandwidths for the vowels /a/ and /o/ whereas the boys tend to have larger bandwidths for the vowel /i/.

Considering the amplitude data, three of the comparisons show boys with significantly higher formant amplitudes than the girls, three show marginal failures of significance and three show no significance. Overall, given any formant for any of the three vowels, the boys either have higher formant amplitudes or there is no difference between the sexes. There is no indication that the girls display higher amplitudes in any case.

These findings seem to support the belief that the girls have a steeper spectral slope than the boys. We know that adult females produce glottal waveforms with lower vocal fold closing velocities, lower a.c. flow and a relatively shorter duration of closure in the closed phase of the glottal period[23] (Holmberg et al., 1988). This combination of physiological circumstances results in the female voice showing a steeper slope in the speech spectrum (spectral tilt) and, as a consequence, causes it to have lower formant amplitudes and wider formant bandwidths. Also, Rabiner and Schafer (1978) showed that the first formant bandwidth is determined largely by vocal tract wall loss and that the second and third formant bandwidths are more dependent on a combination of mechanisms such as wall loss, radiation loss, viscous friction and thermal leakage. The results of Childers and Wu (1991) indicate that the phenomena of wall loss and radiation loss may be larger for female speakers.

The finding that the boy speakers in this study had larger formant amplitudes than the girl speakers is perhaps to be expected even given the differences in spectral slope between the sexes. In the same way that frequency is the acoustic parameter which corresponds to the auditory percept of pitch, so intensity is the acoustic parameter which accords with the sensation of loudness. The intensity of a sample of speech is proportional to the mean amplitude of variations in air pressure (i.e. acoustic energy) and, as vowels tend to have the highest intensity of all speech sounds, they are good markers of individual differences in intensity levels (Ladefoged, 1982). We may tentatively assume that as the boys displayed higher formant amplitudes than the girls this will be reflected in their overall perceived loudness. This may represent a physiologic / aerodynamic cue for gender perception if the listeners responded to those voices which were impressionistically louder than others by marking them as boys. The drawback with this line of reasoning is that it is not possible to relate the measures of formant amplitude presented here unequivocally with perceived loudness as the speakers' distance from the microphone was only kept approximately constant and, whilst every effort was made to keep the recording level static when recording speech from different children, it was occasionally necessary to alter the level to a more suitable point whenever a child spoke with a particularly loud or soft voice[24]. Given these caveats it is possible to conclude only that, amongst the speech data collected, the patterns which emerge

[23] See section 1.1.7

point to two situations: the boys either appear to speak with generally louder voices than the girls or there are no significant differences between the sexes. The different situations arise from the choice of vowel and formant as a measure.

The results of statistical tests between the formant values of the most and least male-like children are listed in tables 3.18-3.20 of section 3.2.1.1 and summarised again in table 4.9. Once again we have the situation where the least male-like children tend to have the higher formant frequencies however on this occasion there are fewer clear-cut cases in which there is statistical significance. More of the individual comparisons have worsened than have become better (signified by superscripted 'w' or 'b' in table 4.9) which indicates that the formant differences between the perceived sex groupings have narrowed with respect to the real sex groupings. There are now only two comparisons with significant differences and the remainder can be interpreted as showing no difference between the most and least male-like children's formant frequencies. The mean *k*-factor of this data, averaged across all vowels and formants is 4.54% however this result must be considered somewhat spurious as there are two means which comprise this grand mean which are several times larger than any other in the analysis - k_2 of /a/ = 12.63% and k_1 of /o/ = 21.59% - and these will result in the grand mean being raised above what would appear to be its more natural level (see footnote 20 for an account of the large sex-difference in the F1 data of /o/).

As with the boy vs. girl analysis, the formant bandwidth data of the most / least male-like children displays much fewer differences. In this case there are three significant bandwidth differences. The least male-like children exhibited higher formant bandwidths for B1 of /a/ and B2 of /o/ whilst the most male-like children had higher bandwidths for B2 of /i/. The remaining comparisons were non-significant but an inspection of the raw bandwidth data reveals that the least male-like children had higher bandwidths in all formants of all three vowels except for B2 and B3 of /i/.

---

[24] It was necessary to ensure that the children's vocal productions fell within certain acoustic limits in order to allow for successful acoustic analysis. If these criteria were met, the inter-child differences in loudness were not tampered with on the recording.

Finally, the formant amplitude comparisons between the perceptually grouped children showed the same patterns as in the earlier analysis but with an even stronger effect. Seven of the nine comparisons showed that the most male-like children had significantly higher formant amplitudes than the least male-like children and the two which failed to reach significance yielded p values less than 0.45. Taken together, the bandwidth and amplitude information again seem to underline the belief that the glottal spectrum of female speakers shows a greater decline (spectral slope) than male speakers

Table 4.9 Results of statistical comparisons of the first three formant frequencies, bandwidths and amplitudes of most and least male-like children's productions of 3 sample vowels. All tests were independent t-tests assuming unequal variances.

	a	i	o
Formant 1	p=0.389 [w]	p=0.367 [w]	p=0.001 [b] *
Formant 2	p=0.032 [w] *	p=0.339 [w]	p=0.214 [w]
Formant 3	p=0.185 [b]	p=0.469 [w]	p=0.423 [b]
Bandwidth 1	p=0.038 [b] *	p=0.400 [b]	p=0.342 [b]
Bandwidth 2	p=0.358 [b]	p=0.016 [b] †	p=0.001 [b] *
Bandwidth 3	p=0.198 [b]	p=0.295 [w]	p=0.440 [w]
Amplitude 1	p=0.003 [b] †	p=0.001 [b] †	p=0.042 [b] †
Amplitude 2	p=0.010 [b] †	p=0.003 [b] †	p=0.419 [w]
Amplitude 3	p=0.200 [w]	p=0.025 [b] †	p=0.001 [b] †

Key

* = Values of least male-like children formant measure > values of most male-like children formant measure
† = Values of most male-like children formant measure > values of least male-like children formant measure
Parentheses indicate marginal failures of statistical significance in the direction indicated (0.1>p>0.05).
[w] indicates that the difference between the groups being compared has lessened from the level in the comparison between boys and girls (table 4.8) - i.e. the value of p has increased.
[b] indicates that the difference between the groups being compared has increased from the level in the comparison between boys and girls (table 4.8) - i.e. the value of p has dropped.

Let us consider this data in relation to the data comparing actual boys and girls.

When we compare the mean formant frequency values of the most and least male-like children with the mean data from the previous calculation of boys' and girls' formants we see that there is a very clear drop in frequency values in the former. The distances (in Hertz) between mean formant values of most / least male-like children and boys / girls are shown in table 4.10. Only six of the eighteen comparisons show negative number results which indicates that in the majority of cases, the mean formant values of the most and least male-like children are lower than those of the actual boys and girls. Furthermore, those results which are negative are very small in absolute terms (no more than 35 Hz) whereas the positive results are much larger - there are two differences in excess of 300 Hz. This would seem to suggest that the children who were identified as being most boy-like and most girl-like both displayed formant frequencies which were lower than those of the average for their respective sexes. It would seem that there is a large amount of high frequency variability associated with the mean formant frequency values of boys and girls which is removed when we consider only those children who are perceptually prototypically male and female. It follows then that those children with very high formant frequency values are poor representatives of their gender, whether they be boys or girls. The largest differences between the boy / girl data and the most and least male-like data are in F1 of /a/ and F2 and F3 of /i/. There remains a formant frequency effect as can be seen from table 4.10. The most male-like children tend to have lower formant frequencies than the least male-like children (see table 4.11).

Table 4.10 Difference between mean formant frequency values of boys / girls and most / least male-like children (Hz). Negative numbers indicate lower formant frequency values for boys / girls.

Vowel	Comparison	F1	F2	F3	Sum
a	Girls / least	321	6	-23	304
	Boys / most	186	136	-20	302
i	Girls / least	0	281	375	656
	Boys / most	-6	202	240	436
o	Girls / least	-24	36	-13	-1
	Boys / most	65	42	-35	72

With the exception of F1 of /o/ in the most / least male-like data, the results in tables 4.8, 4.10 and 4.11 indicate somewhat surprisingly that the largest differences between the sexes (or between the perceived sexes) are in the second formant frequency[25]. The results of Bennett (1981) suggest that the sex-differences in formant frequency increase as the formant number increases.

Table 4.11 Difference between mean formant values of most and least male-like children (Hz). Negative numbers indicate that most male-like children's formants are higher.

	a	i	o
F1	-42	12	109
F2	254	97	83
F3	82	-37	-26

Perhaps a reconsideration of Fant's *k*-factors will shed some light on this subject.

He determined that formant frequency differences between males and females are generally largest for F2 and F3 of the front vowels and for F1 of /æ/ and that they are smallest for F1 and F2 of rounded back vowels and for F1 of any close or highly rounded vowel. For our data, the largest value of k_1 is, as predicted, for the vowel /a/ and, again as Fant anticipated, the smallest value is for the rounded back vowel /o/. For k_2, the value of the front vowels /i/ and /a/ are roughly equal and both are considerably higher than the value for /o/. K_3 involves the highest value for the close front vowel /i/, followed by the open front vowel /a/ and the rounded back vowel /o/ again shows the smallest value. The *k*-factor results can be summarised by formant as follows: F1: /a/ > /i/ > /o/; F2: /a/ ≅ /i/ > /o/; F3: /i/ ≅ /a/ > /o/. They can be summarised by vowel as follows: /i/ and /o/: $\boldsymbol{k_1 \cong k_2 > k_3}$; /a/: $\boldsymbol{k_1 > k_2 > k_3}$.

A glance at figure 3.12 in section 3.2.1.1 (*k*-factor chart) is enough to raise an intriguing question. Why are the male / female differences in formant frequency

[25] There is one suspiciously low F1 value in the most male-like data of /o/ which serves to lower the mean formant frequency of that group. If this stimulus is removed, the mean F1 of the most male-like children approaches that of the least male-like children and the distance figure quoted in table 4.10 is reduced.

relatively smaller in F1 of /i/ than in F1 of /a/? One might continue this line of reasoning and ask why the differences are less pronounced in the higher formants. To explain these findings we must return to the writings of Gunnar Fant. In 1960 he made use of two techniques of theoretical modelling to account for the dependence of formant frequencies on the magnitudes of the resonators which comprise the vocal tract. He used double Helmholtz resonator theory and line analogue measurements to explain how differences in the two main resonating chambers or in lengths of the vocal tract itself could result in changes of formant frequency. Fant claims that the relative value of F1 of /i/ is dependent on both tongue constriction and on pharynx length (Fant, 1960; 1973). We saw in section 1.1.5 that sex-differences in pharyngeal measurements have been reported in adults. Given this finding and the possibility of voluntary movements of the larynx in the vertical plane which could further enhance any sex-difference, we would predict a reasonably large distinction in F1 of /i/ between males and females. As there is not such a large difference, it is likely that the tongue constriction and not the pharynx length is the main determinant of F1 of /i/ in children. Bennett (1981) points out that sex-differences in the degree of mouth opening or in the position of the tongue constriction will not greatly affect the frequency of F1 as long as the tongue constriction is narrow. Furthermore, as the acoustic consequences of pharyngeal differences tend to be reduced when the tongue constriction is narrow, it is not altogether surprising that the differences in F1 of /i/ between males and females are not large. The fact that the sex-differences in formant frequency of /i/ is larger for adults than for children can be interpreted as being due to the relatively larger sex-difference in pharynx length in adults.

There are a number of reasons for the larger sex-difference in F1 of /a/. Firstly, the vowel /a/ does not require any large constrictions of the vocal tract either in front or back cavities and because of this unconstricted passage and also because it is often accompanied by a large degree of mouth opening (Stevens and House, 1955) "the acoustic output [...] reflects the dimensions of the entire tube and as a result provide[s] for maximum distinction between the two sexes on the basis of vocal tract size." (Bennett, 1981: [388]). Secondly, the vowel /a/ is also prone to sex-specific articulatory differences. The vowel is highly robust in that it can allow large variations in the degree of mouth opening and in the size and placement of the greatest point of tongue constriction without loss of its perceptual identity. Lip and jaw opening will, however, affect the frequency of F1, as will larynx lowering. Bennett and Weinberg (1979b) suggest that the lower F1 of males' /æ/ could be partly due to the smaller degree of mouth opening often used by male speakers, whilst Lindblom and Sundberg (1971) have shown that vertical adjustments of the larynx can bring about shifts in F1. Casual observation would also support the idea of a larger sex-difference in F1 of /a/ than of /i/. Perceptually, /i/ tends to vary less across different accents and across different individuals than /a/ - this can be linked to a more consistent tongue position in the former. Overall, there are a number of effects, some organic, others voluntary which will result in a large sex-difference in F1 of /a/ and this is reflected in the values of *k* reported in this study and in those by Fant (1973) and Bennett (1983).

There are a number of conflicting results in the field of *k*-factors. For example, Kent and Forner (1979) state that the F1 scale factor is larger than the F2 and F3 factors in the comparison of adult women with men but that this gradient is maintained only for the vowel /i/ in children[26]. Bennett (1983) on the other hand, showed that the F1 scale factor was larger for all of the vowels which she used **except for /i/ and /ɪ/**. The results of the present study reveal that the mean scale factor of F1 is indeed larger than that of F2, but only marginally so, and the scale factor for F1 of /i/ is, in fact, slightly smaller than that of F2.

The *k*-factor results of the vowel /o/ in the present study are particularly revealing. For all formants the value of *k* is lower for /o/ than for /i/ or /a/. This is partly to be expected as Fant proposed that k_1 and k_2 of the back rounded vowels would be low, particularly so for the highly rounded vowels like /o/. However what is most interesting is that Fant has determined that "the second formant of the back vowels /u/, /o/ and /ɑ/ is somewhat more dependent on the front cavity than on the back cavity" and that "F3 of /u/, /o/ and /ɑ/ is chiefly dependent on the parts in front of the tongue constriction" (Fant, 1960: [121]). He also states that "the contribution to F1

[26] The reader should note that Kent and Forner (1979) use a slightly different formula to derive *k* which involves comparing child formants with formant values of adult males. This will therefore result in much larger values of the scale factor than is seen in this study or in that of Bennett (1983).

of /u/ from the back cavity volume is somewhat larger than that from the front cavity" (Fant, *ibid.*). Interpreting these comments with reference to the present discussion, we can think of F1 of /o/ being dependent on both cavities with slightly more emphasis on the pharynx and F2 and F3 being most affected by the configuration of the oral cavity.

As we have seen, Fant (1973) went on to provide evidence that although there are sex-differences in both front (oral) and back (pharynx) cavities in adults, children probably do not show the same pattern of differences. Fant cited Hunter and Garn (1972) and Walker and Kowalski (1972) in support of his opinion that there are probably no (or only very small) differences in the dimensions of the oral cavities of boys and girls and that the vocal tract length differences could be accounted for very well by considering only pharynx differences. This is another reason why the mean formant scale factors of Bennett (1983) - 10%, and of this study - 3.8%, were considerably lower than those of the studies involving adults (Peterson and Barney, 1952 - 19%; Fant, 1973 - 18%) as the formant frequencies of the adults were varying due to the action of two cavities whilst the children varied according to only one.

As F2 and F3 of /o/ are mostly dependent on the front cavity which displays little or no sex-linked variation, the formant scale factors will be very low. It is interesting that the scale factor for F1 which is the formant which is most dependent on the rear cavity is the highest of the three.

Fant's comments regarding F3 reveal that for each of the vowels used in this experiment, the major determinant of the formant frequency is the part of the vocal tract in front of the tongue constriction. This reliance on the front cavity is borne out by the present formant scale factor results. It is clear that the values of k_3 for all vowels are substantially lower than those of k_1 and k_2 and they reflect the distinction between the unvarying front cavity and varying back cavity. Within k_3 the value for /i/ is highest, reflecting the fact that, in addition to the front cavity contribution, the formant value relies on either some degree of tongue constriction information or some back cavity input. The lowest value of k_3 is for the vowel /o/. As Fant puts it, "in [...] /o/ F3 is associated with a three-quarter wavelength resonance of the cavity system in front of the tongue constriction" (Fant, *ibid.*).

The largest differences in mean formant frequency between the boys and girls in the present study are to be found in the values of F2, especially in the vowels /i/ and /a/.This is in contrast with Bennett (1983) who found all formant differences to increase as the formant number increases. According to Fant's calculations using Helmholtz double Resonator theory, F2 is highly influenced by tongue advancement. It is possible, therefore, that the girls in the present study are producing their vowels with a greater degree of tongue advancement within the limits required to maintain phonemic identity. This effect is greatest for the front vowels /i/ and /a/ indicating that the girls are aiming for more peripheral (extreme) targets than the boys. This situation is not without its linguistic precedents. Trudgill (1983, [78]) reports that "in many accents of American English it has been found that women's vowels are more peripheral [...] than men's". Trudgill goes on to point out a number of examples of linguistic phenomena in which female speakers consistently use forms which are perceived as "better" than the male variants and states that women "use forms which more closely approach those of the standard variety or the prestige accent" to a far greater degree than men (Trudgill, *ibid.* [85]).

In sociolinguistic terms this finding is closely related to the concept of linguistic conservation / innovation. In this case, as with many other examples, the females are seen as preserving the established form of the production (the more peripheral vowels) whereas the males prefer the 'newer', often reduced forms (the more central variants). The reasons behind such phonetic variations may be traced back to the notion of prestige (Trudgill, *ibid*). There are a number of phonetic, phonological and syntactic constructions which often vary depending on social class of the speaker, for example, higher class speakers use fewer cases of non-standard multiple negation than lower class speakers. In broad sociological terms, women in Western society tend to be more status-conscious than men (Eakins and Eakins, 1978). Therefore women tend to be more aware of the social significance of these social class-related linguistic constructions. It is also the case that lower-class culture, including the speech characteristics of lower-class speakers, is often associated with masculinity and male speakers will therefore be more inclined to adopt speech patterns which display non-standard varieties than female speakers. The connection between lower-class speech and the 'toughness' which is characteristic of lower-class life may lead to the link with masculinity, which is perceived as an advantageous masculine trait.

Trudgill summarises the situation as follows,

> "Given that there are linguistic variables which are involved in a speech community, in co-variation with social class (higher-class forms being more statusful or 'correct' than lower class forms), then there are social pressures on speakers to acquire prestige or to appear 'correct' by employing the higher-class forms. Other things being equal, however, these pressures will be stronger on women, because of their greater status-consciousness. On the other hand, there will also be pressure [...] to continue using less prestigious, non-standard variants as a signal of group solidarity and personal identity. These pressures, however, will be stronger on men than on women, because of concepts of masculinity current in our society. *Men's speech will therefore be less 'correct' than that of women.*"
>
> (Trudgill, 1983, [87-88]: my italics)

This issue is relevant to the 4-5 year old children in the present study although we might prefer to focus more directly on the concept of gender role than on social class. Arguably, children learn to adopt gender-specific behaviour (and therefore adopt *a gender role*) gradually over the first few years of life. Although much has been written on gender roles in adults and the effect of gender on patterns of socialisation in the community at large, what is sometimes forgotten is that pressure to conform to gender-appropriate behaviour is probably greatest in the years prior to puberty. During adulthood there is a certain amount of freedom to express one's gender role and / or sexuality in various ways - not all of which may be in keeping with the 'default model' prescribed by normal pressures of socialisation. It is relatively straightforward, for example, for a man to make use of characteristically feminine behavioural traits and to display aspects of his personality which we might more usually associate with females, just as it possible to encounter women who favour generally aggressive, masculine behaviour and follow value systems which are usually adopted by males. On the other hand, there is considerably more pressure, for example, on a 7 year old boy to associate with his friends and to play football with them and engage in 'rough and tumble' play activities. His behaviour would not be tolerated if he chose to ally himself totally to a group consisting of girls and to shun the typically masculine-orientated manners of his male peers in favour of what would be perceived as the weaker, less-valued, feminine-orientated behaviour of the girls. The social pressures acting on these children, in particular the boys, are powerful indeed and function to constrict and channel the possible behaviours into only those which have been 'selected' as suitable for the particular gender.

Despite the findings that formant frequencies, bandwidths and amplitudes are all involved to some degree in the gender identification process, the actual importance of each remains unclear. The multiple regression analyses carried out indicate that the combined formant frequencies were able to account for around 54% of the variance in the perceptual judgements of maleness and the bandwidths and amplitudes each contributed relatively less. It seems most likely therefore, that the listeners used no single acoustic cue which supplied very high levels of gender information but combined the available cues in such a way as to maximise the contribution of each. The precise nature of this mental process remains unclear, however a recent study has examined the perceptual representation of talker voice and specifically voice gender and has concluded that,

> "...these representations are not based on one isolated parameter like Fo or formant frequencies. Instead, the representations are probably an auditory composite of the various acoustic factors relevant to voice gender like Fo, formant frequencies, breathiness, etc.".
>
> (Mullenix, Johnson, Topcu-Durgun and Farnsworth, 1995: 3091)

4.2.1.2 Breathiness measures

There are a number of difficulties inherent in the assessment of vocal breathiness. Firstly, it is known that there are many spectral, acoustic and physiological correlates of the auditory effect of a breathy voice. For example, the use of relatively low vocal fold tension in conjunction with high airflow rate as in /h/, or the adoption of a vocal fold configuration which allows a large posterior glottal chink whilst maintaining normal vibratory adduction at the front of the folds as in the phonologically breathy segments of languages such as Hindi or Gujarati (see section 1.1.7.1 for a fuller account of breathiness).

The measures adopted in the present study to assess breathiness follow Bickley (1982) and Henton and Bladon (1985) in that they represent the absolute differences[27] in amplitude between the fundamental component and the second harmonic (H1-H2) or the first formant (H1-F1). Temple (1988) discusses the implications of reporting the breathiness measures in different ways. She points out that in the analysis of the voices of speakers with widely different Fo, such as adult males and females, the same absolute amplitude measures can be associated with different spectral slope measures. As the present study is concerned only with pre-pubertal children who display similar Fo values, the use of absolute amplitude measures is valid.

Due to the relatively high fundamental frequencies of the child speakers, there will be an interaction between Fo and F1 in all but the phonetically open vowels. The open vowel /a/ has a high F1 and therefore should be the vowel whose formant structure affects the lower frequency harmonic components to the least extent. Without inverse filtering the data, which is often a complex and unreliable procedure and, in any case, was not available for this study, it is not possible to negate the effect of the transfer function on the glottal waveform. Henton and Bladon (*ibid.*) restricted their experiment to the vowels /æ/, /ɑ/, /ʌ/ and /ɒ/ specifically in order to avoid source / filter interaction in their measures of breathiness.

The discussion of the current breathiness measures will therefore be restricted to the results for the vowel /a/.

The finding that there are no significant differences in either measure of breathiness used in this study between boys and girls is of considerable interest. In their study of breathiness in adult British speakers, Henton and Bladon (1985) showed that females

[27] The reader should note that although the term "absolute difference" is used with reference to H1-H2 and H1-F1 measurements in order to emphasize that these are not direct spectral slope measures, the decibel is, of course, a relative unit of amplitude.

spoke with significantly breathier voices than males and that this phenomenon was consistent over different accents and across a number of open vowels. When we compare the findings of that experiment (H1 - H2: vowel /æ/ - females 8.4 dB, males 0.98 dB) with the results of the present study (H1 - H2: vowel /a/ - females 5.48 dB, males 5.85 dB) we can see that, even allowing for accent differences in the speakers used in the two experiments, the values of H1-H2 of our child speakers fall roughly between the values of the adult males and females of Henton and Bladon's (*ibid.*) experiment. A number of researchers (Klatt and Klatt, 1990; Henton and Bladon, 1985; Ladefoged and Antoñanzas-Barroso, 1985) have pointed out that the increased use of breathiness in adult female speakers of English (where there are no phonemic contrasts signalled by breathy phonation) is probably a social marker of gender. If breathiness is used as a cue to femaleness, the question arises, when do girls / women acquire or begin to employ it? If we are to believe the results cited above then clearly by the time of maturity there is a contrast which is not present at the ages of 4-5 years given the present results. It would makes sense to assume that fine control over breathiness is acquired after the ability to manipulate all of the phonetic parameters involved in the linguistic formation of phonological contrasts. In other words, a child must learn the distinctive features of his or her language in order to recognise that other phonetic features can be used paralinguistically, not to signal differences in utterance meaning but rather to relay indexical information.

The subject of age-related differences in vocal breathiness appears to be poorly represented in the research literature. As there are no indications of the development of breathiness other than as a part of the general growth in neuro-muscular control and refinement of phonatory and articulatory patterns which are co-requisites of normal speech development, it is possible to speculate on the evolution of the adult contrast in breathiness.

The values of H1-H2 discovered in this study for the prepubertal children (around 5.5 dB) are large enough to reflect a certain amount of breathiness in the voices of both boys and girls. Although it was not specifically required of them, none of the listeners in the present study made a comment as to the breathiness of the children's voices, nor did phonetically trained adults who informally heard extracts of the recordings. Temple (1988) quotes Watson (1987) in a discussion of the perception of breathiness, "It may be that we accept as normal in children what would be 'breathy' in adults, until we are specifically called on as phoneticians to attend to phonation type" (*ibid.*, p.21). The implication here is that the phenomenon of breathiness is a relative one and can only be assessed with reference to a normative level for modal phonation, or, in the case of sex-differences, with respect to a sex-specific (and culture-specific) norm. In the same way that the voice of a prepubertal child is more like that of an adult female by virtue of the closer Fo levels, so the degree of breathiness appears to be closer to that of the woman than the man. We can imagine that with the greater growth of the male vocal folds in puberty and the general enlargement of the volume of the larynx, these physiological explanations for greater and stronger vocal fold contact behaviour in adult males would tend to lead to a reduction in the amount of vocal breathiness. In the female there is much less of an increase in vocal fold mass which if combined with learned voluntary adjustments of the mode of phonation could easily account for the habitual breathiness seen in many adult female native speakers of English.

The second measure of breathiness (H1-F1) compared the amplitudes of the fundamental and first formant components of the vocal spectrum. As the results for the vowel [a] were all less than zero and as the mean H1-F1 values were around -14 dB, it is clear that there was considerably more acoustic energy at the first formant frequency location than at the fundamental. Ladefoged and Antoñanzas-Barroso (1985) measured modal and breathy vowels[28] in speakers of !Xóõ and although they discovered that breathiness could only be characterised in relative terms there were clear trends in the H1-F1 and H1-H2 values of the two different phonation types. Table 4.12 compares mean spectral measures of breathiness in the Ladefoged and Antoñanzas-Barroso (*ibid.*) study and in the present study.

[28] The authors report that the vowel samples were taken from the words //áa (a camelthorn tree) and !äö (a slope). Although there was no transcription provided with the glosses, it is assumed that the vowels had qualities broadly similar to [a].

Table 4.12 Comparison of breathiness measures across two studies.

Spectral measure	Ladefoged and Antoñanzas-Barroso		This study
	breathy vowels	modal vowels	
H1-H2 (dB)	13.6	-7.5	5.7
H1-F1 (dB)	-4.9	-28.1	-14

The pattern of results, given that there are only two studies for comparison, is interesting. There are three dimensions which we can compare: firstly, the distinction between the breathy and modal vowels in Ladefoged and Antoñanzas-Barroso's study. In both spectral measures, the distinction is characterised by the breathy vowels showing higher values of the measure by approximately 20 dB. This is a reflection of the increased energy centred on the fundamental component of breathy vowels - larger values of H1 (amplitude of fundamental component) results in larger values of H1-H2 as discussed in section 1.1.7.1. Secondly, the difference between the two measures of breathiness is also characterised by a difference of about 20 dB, with H1-H2 values being larger than H1-F1 values. This situation reflects the relatively large degree of acoustic energy localised at the first formant. As breathiness appears to be a relative measure (in terms of the spectral measures examined here at any rate) there is no problem in stating that a value of H1-F1 of -4.9 dB represents a breathy vowel because, although it is a negative figure and falls closer to the value of H1-H2 of the *modal* vowels, it is still around 20 dB greater than the H1-F1 result which represents the modal vowels.

Finally, the breathiness results of the present study fall roughly halfway between the values of the breathy and modal vowels of Ladefoged and Antoñanzas-Barroso. It is difficult to interpret the status of a single breathiness figure given the relative nature of the feature as we have just discussed. However, as the 20 dB difference between the two different spectral measures is clearly shown, it is perhaps reasonable to tentatively compare the results across the two studies. As stated, the breathiness measures of this study fall between the figures quoted for breathy and modal vowels in Ladefoged and Antoñanzas-Barroso's study, however in both cases the figure is closer to the level of the breathy vowel than to that of the modal vowel. This suggests that the present children's vowels are neither convincingly breathy nor modal but tend towards the breathy end of the spectrum - a situation which is consistent with the finding that the breathiness results fall between the values of adult men and women but, overall, the voices of the prepubertal children are more female-like than male-like.

Figure 3.17a in section 3.2.1.2 shows the distribution of the children's H1-F1 values of [a] plotted against the perceptual judgements of their gender. There is a clear downward trend in the data of both boys and girls indicating that those children whom the listeners judged to be most male-like had the lowest values of H1-F1 (i.e. were least breathy).

If the effect was a consistent one, we would expect that the variation around the trend line would be approximately equal across the whole distribution from least male-like to most male-like. However, it seems that there is an unexpected range of H1-F1 values at the extreme male-like end of the perceptual spectrum. In other words, whereas we would expect that the H1-F1 values should be clustered around a minimum value where 'maleness' is high, the actual values of H1-F1 are spread over a large range of values (roughly from 0 to -25 dB). When we remove the data from those children with maleness ratings of $\geq$ 15, the trend lines on the graph show a better fit to the data, indicating that the breathiness values of the most male-like children are unrepresentative of the pattern of the remainder of the data (see figure 3.18a, section 3.2.1.2). Surprisingly, when the other spectral measure of breathiness (H1-H2) was plotted against perceptual judgements there was no correlation whatsoever.

Looking at this effect from the listener's viewpoint, it seems that listeners can use at least one index of spectral slope (H1-F1) to distinguish between the most male-like and least male-like voices for the majority of children. This parameter however does not seem to be useful (or listeners choose to make use of another perhaps more reliable cue) in the identification of the most clearly male-like voices. Perhaps it is the case that breathiness as measured by H1-F1 is a reasonable cue to the gender of otherwise perceptually ambivalent children but is less successful in comparison with the cues which signal very well identified children.

Once again the reader should be aware of the relatively large amount of variability displayed by children in terms of their breathiness values. Although the removal of the data at the extreme 'male-like' end of the range improves the overall variability, the coefficients of regression which quantify the goodness of fit of the trendlines do not exceed $R^2 = 0.32$. Furthermore, although the boys' data is found generally on the right of the graph where the maleness scores are large and the girls' data is similarly situated to the left side of the graph, there is considerable overlap in the centre, indicating that some of the most male-like children are girls and some of the least male-like children are boys.

The issue of why the measure H1-F1 functions to separate the most male-like from the least male-like children better than it does the actual boys and girls is best appreciated by examining figure 3.17a, section 3.2.1.2. If we concentrate on the boys' and girls' data first (i.e. the squares and diamonds in the chart) it is clear that there is some (but by no means a perfect) division of the children into sex groups brought about by the maleness ratings. In particular, there are a number of boys (squares on the chart) whose data points fall in the midst of the girls' points at the low end of the maleness ratings. Overall, the boys' H1-F1 values are not significantly lower than the girls' values - a fact which is partly due to the large variability seen in all of the data. Considering now the most and least male-like children's data, there is a perceptible slope from upper left to lower right on the chart showing that maleness and H1-F1 are approximately inversely proportional. How can this be the case if the boys' and girls' H1-F1 data was roughly equal? The answer lies in the distribution of H1-F1 values <u>within</u> each sex grouping. The least male-like girls tended to have the highest H1-F1 values whereas the most male-like girls had lower values. Similarly amongst the boys, the least male-like boys had higher values of the measure than the most male-like boys (with the exception of the very best identified boys as discussed above). The outcome of this situation therefore is two sets of H1-F1 data (one for boys and one for girls), both of which decline as we climb the maleness scale but which start from roughly similar positions. In fact, we can see from figure 3.16 and table 3.26 of section 3.2.1.2 that the girls do have a slightly higher mean value of H1-F1 than the boys.

To summarise, the girls' H1-F1 data declines over a range of values centred around -13.39 dB and the boys' data declines over a range of values centred around -14.49 dB. When the two sets of data are grouped together they form a larger set which appears to behave according to one general slope although we can identify two separate but similar patterns of distribution within the set.

4.2.2 Sentences

The acoustic parameters extracted from the sentence samples fall into three categories. There appear to be surprisingly few acoustic differences between the boys and girls, on the basis of these statistics.

Fundamental Frequency Information

The data comprises averages taken across 2 different sentences and spoken by 89 different children yet we see, for example, that the average Fo values are 263.15 Hz for girls and 263.78 Hz for boys. The Fo ranges are also identical (within ± 1Hz).

One of the more noticeable differences is in the standard deviations of Fhi (the highest Fo value used). This was 58.6 Hz for girls and 36.7 Hz for boys. This indicates that the girls tended to be slightly more variable about the upper limit of their Fo range although the mean figure is equivalent to the male figure.

Given that the majority of studies discussed in section 1.2.2 found no significant differences between the fundamental frequencies of pre-pubertal boys and girls, and that those which did find small differences were generally dealing with older children than are involved in the present study, it is not altogether surprising that there is such an overlap in Fo values between the sexes in the current results.

We saw in section 1.2.1 that there was some evidence to support the quantal theory of vocal fold growth in children prior to puberty and that this theory necessarily entails disproportionate patterns of growth in boys and girls. However, we also noted that the greatest period of disproportionate growth was at the pubertal growth spurt itself and, prior to this point the growth rates of boys and girls appear to be nearly linear and roughly equal. Given this circumstance and the fact that there is considerable vocal ligament growth from the age of around 3 years which brings about an increase of stiffness in the vocal folds and tends to cancel any effect of a vocal fold length increase on the mean Fo, the available evidence clearly suggests that we should find little, if any, fundamental frequency differences between the sexes at this age.

Short and Long Term Frequency and Amplitude Perturbation Information

As discussed in section 2.2.2, the values for both boys' and girls' frequency perturbation measures (jitter) and amplitude perturbation measures (shimmer) were considerably higher than are reported for adult subjects in the research literature. Given that the measurements of jitter and shimmer were taken over a large time-domain and therefore represent long-term averages, this is to be expected. The pitch varying influences of intonational movements and stress placement will combine with the random components which represent the true frequency and amplitude perturbation elements of interest to yield an overall 'yardstick' measurement which we can compare between the sexes. In this case, the fact that there were no significant differences between the jitter or shimmer values of boys and girls is perhaps not very surprising.

The reader should bear in mind that jitter and shimmer are primarily intended to measure the regularity (or irregularity) of the vibrations of the vocal folds. As such they are good predictors of certain vocal pathologies. However, as all of the children in the study were selected to be free of speech and hearing deficits and therefore had no laryngeal problems, the finding that the boys and girls do not differ in terms of these measures is unremarkable. Secondly, although shimmer is thought to be as important as jitter in the contribution to the perception of vocal hoarseness, a condition seen to be more prevalent in young boys than in girls, "The effect on amplitude perturbation of neither fundamental frequency nor mean amplitude has been explored...*Possible differences due to age or sex remain unclear.*" (Baken and

Daniloff, 1987 [my italics]). Baken and Daniloff (*ibid.*) also quote shimmer values for adult males and females which indicate larger values for males, although these are vowel-specific and non-significant. Other researchers have reported that shimmer separates normal speakers from pathological speakers far more accurately than it does normal males from females (Davis, 1979; Takahashi and Koike, 1975).

Jitter values appear to correlate with fundamental frequency in adult speakers - larger frequency perturbations are associated with lower Fo (Horii, 1979, 1980). Adult males therefore tend to have higher mean jitter values than adult females. This situation accounts for the equality of the jitter measurements in the speech of the prepubertal children in this study. As there is no mean Fo difference between the boys and girls, and, as neither group is comprised of speakers with laryngeal pathologies, there is no significant difference between the mean jitter values of the sexes.

4.3 GENERAL SUMMARY AND CONCLUSIONS

There were three main questions which motivated this research at the outset:

- Is there a measurable success on the part of adult listeners to identify the gender of prepubertal children on the basis of voice alone?
- If so, what can we say about the impact of differences in child and adult gender on the recognition process?
- To what extent are different cues involved in the process of gender identification?

We are now in a position to relate the findings of the study to these basic questions and to comment on issues that have arisen throughout the research as a whole.

The first two questions, which relate to the ability of the adult listeners, were dealt with in the perceptual side of the study. There clearly is a demonstrable ability to recognise gender. On average, judges could identify boys and girls from voice alone at a rate which was better-than-chance. A secondary issue of the gender identification rates concerned potential differences in judgement accuracy across a variety of speech sample types. It was established that the two connected speech sample types used in the experiment (sentences and quasi-spontaneous speech) resulted in better rates of gender identification than the isolated vowel sample. Two hypotheses were suggested to account for this finding. The first focused on the fact that the sentence and spontaneous speech samples contain more information than the vowels whilst the second highlighted the fact that there is a qualitatively different type of information in the connected speech which might contribute relatively more to the gender identification process than the isolated vowel information.

It also became apparent that there were effects of both child sex and judge sex operating in the experiment. Overall, girls were better identified than boys by all judges. When this finding is combined with the result that the female judges tended to be better at the task than the male judges, we can begin to recognise a pattern of results which reminds us of the widespread idea of the superiority of girls with respect to almost all aspects of language processing (Templin, 1957; McCarthy, 1953). It is not clear how the finding that girls are better identified than boys relates to the general linguistic advantage of females, however factors such as earlier maturation of verbal fluency and adult-like competence may be involved. Acknowledging a sex-difference in favour of females is one thing; accounting for it is another more complex task. There are a number of different levels on which researchers attempt to explain sex-differences and, although some of these are more appropriate than others for the purposes of rationalising children's speech sound differences, it is most likely that a number of these factors influence the sexes differently.

Biological / Physical differences. A logical starting point in the examination of sex-differences is to assess to what extent the variation between males and females is innate. It is reasonable to suggest that the different anatomical structure of males and females might give rise to different communicative outputs in terms of certain relevant vocal parameters. However, as we have seen, whilst the differences that exist between the vocal organs of fully developed men and women are usually sufficient to allow listeners to distinguish speakers according to sex, this is not the case with immature (pre-pubertal) speakers. In chapter 1 we considered two possible developmental paths for vocal fold growth: quantal and linear. Whichever of these paths is the more accurate reflection of the children's pattern of development, neither predicts a sufficiently large distinction in vocal fold growth by the age of 5 years (the average age of the subjects in this study) to account for a perceptual distinction. This is reflected in the mean fundamental frequencies for the boys and girls which were found to be almost identical.

The other major perceptual difference between adult males and females is the formant frequencies which depend, to a large extent, on the vocal tract length. The formant measurements taken in this study and others indicate that there is a small difference in average formant frequency between boys and girls and that taken as a whole the formants are a far better cue to gender than the fundamental frequency. Statistics suggested that around 54% of the variance in the gender perception data could be accounted for by the combined formant information. Whilst this represents a substantial degree, it certainly does not fully describe the gender recognition ability of the judges. Furthermore, without taking specific vocal tract measurements (a feat which is extremely difficult to accomplish with any accuracy on living subjects) we cannot be sure as to whether the formant differences we see are due to innate factors or to other voluntary manipulations of the vocal organs (see *cultural influences* below).

Personality. Another possibility is that males and females may differ innately in personality. Sociologists argue as to the extent that an individual's personality is dictated by the hormonal and chemical make-up of the body (innate) and the experience and interaction that the individual has with his or her environment (learned). In reality it is almost certainly the case that these two factors interact to form a structured personality - we may occasionally act on biological impulses, however cultural and social pressures probably override many innate influences and mould us into socially acceptable creatures. Eakins and Eakins (1978) suggest that if there are innate personality differences between the sexes, they would probably result in behaviours which overlapped to such an extent that it would be impossible to differentiate the personality traits by sex.

In sociological terms, the involvement of personality in the construction of sex-differences is an essentialist approach in which gender is seen as a fundamental, essential part of the individual (Crawford, 1995). According to this view, gender is seen as a set of properties which comprise one's personality. Gender is something women and men *have* or *are* - it is a noun (Bohan, 1993). Therefore the dual findings that women tend to lack the confidence and specific skills to speak assertively in public, but do value intimate, co-operative discourse strategies are both based on an essentialist view. "Women speak in particular ways *because they are women*" (Crawford, 1995; 8). There is nothing in essentialism which specifically implies a biological origin for gender-specific features, indeed, the emphasis tends to be on the location of characteristics within the individual and not their origins. It is these stable, influential characteristics (which include masculinity or femininity) which are thought to determine gendered identities and behaviour.

Cultural influences. In contrast to essentialism, the theory of social constructionism sees gender as a meaningful filter through which individuals interact and can access power and resources. In this view, gender is not an attribute of individuals, but of the actions and interactions of the individuals. "It [gender] is conceptualised as a verb, not a noun" (Crawford, 1995; 12).

Anthropologist Ray Birdwhistell (1970) studied patterns of non-verbal behaviour and concluded that sex-differences in gesture, expression, posture and body movement are socially motivated and culturally learned. There is certainly also a certain degree of cultural influence on speech patterns. Fant (1966) and Mattingly (1966) both pointed out that formant frequency measurements which they had taken of adult males and females were more widely separated than one would anticipate if the sole distinguishing feature was the different anatomies of the sexes. These researchers chose to use the vowel /æ/ because of its particular ability to show up individual differences in articulation habits. The importance of the vowel stems from the fact that it is possible to pronounce a recognisable /æ/ with a variety of tongue positions and constrictions and different degrees of mouth opening. Bennett and Weinberg (1979a) suggest that, for their subjects, "it is possible that sex-specific variations in the degree of jaw opening and the amount of lip rounding associated with the production of /æ/ also provided relevant information about sexual identity".

In the present study, a number of formant frequency differences were found between the prepubertal boys and girls. If we assume that there is a negligible difference in vocal tract size between the sexes prior to puberty[29] then the phenomenon of voluntary adjustments of articulations in order to conform to cultural stereotypes can be extended from the adult situation, where it enhances the already present biological differences, to the situation of the prepubertal child, where it largely explains the differences between the sexes. It also provides a partial explanation of the fact that not all prepubertal girls have higher formant frequencies than prepubertal boys although the majority do. The rationalisation is that if the formant differences are largely dependent on the individual child adopting a particular mode of articulation which he or she has learned is appropriate for his or her gender, then different children will acquire this feature at slightly different times. For example, if a girl is slightly slower in adopting this sex-specific habit than her peers, then her formant frequencies will be slightly lower than those of her female peers who have acquired the practice and, in fact, more characteristic of the boys of her age. Perhaps it is the boys who learn to use particular speech habits, for example, lip rounding, which would tend to lower their formant frequencies; or perhaps the sexes simultaneously adopt different patterns which cause their formants to diverge. Whatever the specific details of the process, this is the general situation which is revealed in our data. We see that in general the girls have higher formant frequencies than the boys, but there are some examples of individuals with formant patterns more characteristic of the opposite sex. It is expected that after a few more years of linguistic and social development, all of these 4-5 year old children will be speaking in ways which much more clearly demarcate them as male or female.

It is also interesting to note that, as with Fant (1966), Mattingly (1966) and others, in our study the vowel which best separated the sexes according to formant frequency was /a/. F1, F2 and F3 of /a/ were all significantly correlated with the degree of perceptual maleness as judged by the listeners. This is further confirmation that there is a process other than the simple scaling effect of vocal tract size in operation here.

We have suggested a possible effect of socialisation on the vocal output of adults and children to account for formant differences. The question of why humans use these mechanisms at all has been skimmed over.

In comparison with certain other species, human males and females do not differ greatly in secondary sexual characteristics; i.e. they are only weakly sexually dimorphic. Social anthropologists suggest that it is directly because of this unimorphism (the relative sameness of males and females) that humans have evolved an elaborate system of gender display. These displays take the form of certain ritualised patterns of verbal and non-verbal behaviour that 'belong to' and mark each sex. It is precisely because the innate differences between the sexes are not that great to begin with that the system of gender display has developed in order to emphasise that which is already present. Seen in this light, the vocal manipulations performed by adults to make males sound more male-like (e.g. larger and more aggressive) and females sound more female-like (e.g. smaller and submissive) fit perfectly into this role of emphasising existing, but relatively small, sex-differences.

It is highly likely that the actual perceptual targets of these manipulations are defined by the social environment in which we live. The observation that females tend to adopt patterns of speech which reinforce the stereotype of a relatively small, quiet, submissive individual may be due to the situation that in our society females are supposedly brought up to be passive, docile, self-effacing and self-deprecating (Lakoff, 1978; Coates, 1995). In the same way, males are acting on social pressures which mould them into positions of dominance, power and prestige. By being permitted stronger means of expressing themselves thanks to their social dominance, males receive reinforcement of their position of strength and it is a natural consequence of the dynamics of communication that listeners tend to pay more attention and give more respect to those individuals who are able to speak out in the

[29] This is an issue which urgently requires elucidation by the gathering of systematic normative data. Until it is established that there are no vocal tract size differences between prepubertal boys and girls, the theory of voluntary, sex-specific articulations must remain only a hypothesis.

most convincing and confident way. It is not that females are incapable of speaking out with confidence and conviction as is clear from highly influential and dominant female speakers (e.g. Margaret Thatcher, Hilary Clinton), however these individuals are heavily outnumbered by their male counterparts. It seems that females may often be 'prevented' from outspoken expression by cultural programming in 'appropriate' women's speech. As psychologist Pamela Butler (1976: 6) states, "The acceptance of traditional femininity clearly interferes with female assertiveness". Therefore, once established, the social roles to which male and female speakers conform are largely self-supporting. It requires a large and sustained effort against the established flow of social interactions to alter social stereotypes.[30]

Despite the difficulties involved in such a process, there have been many changes to the perceptions of the traditional sex-roles in the past decades. Our society delegates certain activities, responsibilities and privileges to people on the basis of sex. Despite the advances in equal opportunities and the new attitudes encouraged by the growth of the women's movement, there is still an overwhelming assumption that childcare and domestic tasks are predominantly 'woman's work'. It is an indication of how persistent certain stereotypes can be when we realise that there are more women in paid employment than ever before and that many households are now ruled over by females and yet traditional notions of the division of labour endure. The issue of sex-typecasting in jobs is not irrelevant to the point of gender differences in communication. There are still relatively few women working in engineering, accountancy or industrial management and similarly few men who work as secretaries, personal assistants or speech and language therapists despite the fact that there is nothing in any of these occupations which precludes (or should even disadvantage) either sex. Even in those occupations in which males and females both work, there is an unequal balance of the sexes across different positions in the hierarchy of the workforce - for example, there are many more senior male consultants in the health services and many more female nurses. There are

---

[30] See Crawford (1995) for a comprehensive discussion of assertiveness in language. Specifically, she argues that psychology and culture have interacted in a complex way to promote the social construct of assertiveness and that it has become a self-fulfilling prophecy for many women who now suffer from learned helplessness. She claims that women have been "encouraged [...] to internalize

undoubtedly communication differences associated with these sex-differences in employment. These differences are probably concentrated at the verbal level of communication, e.g. the different duties carried out by men and women may lead to differing conversational topics, dissimilar styles of talk, distinctive vocabularies and also divergent patterns of non-verbal communication. Furthermore, the differences in employment patterns between the sexes is partly brought about by the dominance of males in society in general and, as with other aspects of cultural moulding, also serves to heighten the power relationship.

A fuller discussion of the influences of social pressures on gender roles is clearly beyond the scope of this thesis. Nevertheless, even a simple understanding of the fact that males tend to have active and instrumental attitudes which relate to their view of the world as something which requires conquest or manipulation through competitive skills and resources, whilst females tend to have much more passive, affective attitudes related to their view of the world as something requiring cultivation through nurture and co-operation, tells us a great deal about why we go to the lengths that we do to develop distinctive gender displays. Although many of these issues await the 4-5 year old children of our study in later life, their effect may already have begun to take root. They may not understand why or be consciously aware of how they are doing it, but they have begun to adopt patterns of behaviour which mark them as being as clearly a member of a social gender role as their genetic make-up identifies them as belonging to a biological sex.

These issues can be related to the findings of the present study. The parameter of formant amplitude is a candidate for a feature which has been shaped by social, psychological and personality influences. As formant amplitude reflects overall vocal loudness to a degree, we might expect that the more socially powerful males with their characteristic aggressive natures would display voices with greater average intensities than the females who have begun to adopt patterns of behaviour which advocate more gentle and docile speech levels. This seems to tally with the findings

---

the image of themselves as passive, inept communicators and to find a cure for this invented deficiency through psychology." (Crawford, 1995 [84]) although the issue of formant amplitude was not completely clear-cut in that the boys did not show higher amplitudes in 100% of cases.

The third question which we posed at the outset of this section related to the cues used by the sexes to convey relevant information regarding gender. We have already considered the importance of manipulation of the formant frequencies, we must now turn, briefly, to the question of breathiness.

It would seem that, in non-pathological, English-speaking adults at least, breathiness is a vocal feature which is either wholly acquired or, as in the case of formant differences, superimposed over and above the small innate influence of gender. The facts are plain enough - in adults, most females are significantly breathier than males; in 4-5 year old children, there are no significant differences in degree of breathiness. If this characteristic is biologically based it is surprising that there is no indication of even a small sex-difference prior to puberty. Much more likely is that there is only a small anatomical influence which causes females to be breathy but that they adopt the property of breathiness as a further cue to gender. This then would be another clear case of cultural elaboration in operation. A breathy voice is associated with weakness, submission and sometimes indicates sexual arousal (Henton and Bladon, 1985). However politically incorrect this may be, these first two traits are more typically linked with females than males. The issue of sexual arousal is a clue that the voluntary adoption of such a feature has an evolutionary significance in addition to pure gender identification information. Whether females adopt breathiness as a strategy to attract the opposite sex is a contentious issue and is discussed elsewhere (Henton and Bladon, *ibid*), however it seems that it is used as a cue to female gender. Adult males display far lower levels of breathiness than females during normal conversation but when they attempt to imitate stereotypical 'female speech' their degree of vocal breathiness increases. The finding of the present study that breathiness levels in prepubertal children were not only equal across the sexes but also roughly midway between the adult male and female values is revealing. It is reasonable to suggest that at an age when the vocal organs are still immature and the mode of phonation has not fully settled into the adult pattern, there will be a relatively large degree of breathiness apparent in the voices of young prepubertal children. By the time puberty has occurred, the vocal folds of both males and females will have grown in length and mass and will be in a better position to make fully complete and abrupt closures - a prerequisite of non-breathy speech. The growth in the males will be greater than in the females and therefore we would anticipate less breathiness due to physiological factors. At the same time, females learn to adopt a weaker (or incomplete) vocal fold closure which results in a breathy output. This suggestion would predict that from a position of equality with respect to breathiness prior to puberty, males develop in a way which reduces the breathiness of their speech whilst females develop increased breathiness due primarily to the acquisition of a learned, gender-linked feature. The breathiness data collected in the present study conforms to this hypothesis.

We have seen that in some ways our voices conform to the stereotypes established by society. Adult women *do* tend to have higher pitched, softer, more pitch-dynamic and breathier voices than men and each of these factors contributes, to a certain degree, to the perceptual ability of listeners to determine their gender from voice alone. Perhaps the most interesting finding of this study is that when we shift our attention to prepubertal children, the perceptual ability to recognise gender remains more or less intact (however with a certain reduction in accuracy) but the obvious sex-specific vocal cues which are so apparent in adults are almost completely lacking in the children. This leads us to two possible conclusions: that the gender cues used by children are different from those which are used by adults; or, that adult listeners are able to successfully establish the sex of the speaker from the tiny source of gender-bearing information available. It is possible that these two conclusions may interact. For example, Fo information appears not to play a part in the perceptual separation into gendered groups with prepubertal children, however formant frequency differences are involved, albeit to a lesser extent than in adults. It may be the case therefore that, as they are already armed with the knowledge that the speakers are young children, listeners are able to mentally scale the formant differences up to a level at which they relate to adult differences or, to state the suggestion differently, that listeners focus their perceptual attention on a much narrower range of potential formant differences and that this narrowing of the field of perception allows for an equalisation of the discriminatory ability with respect to adult speech samples.

Earlier, we considered the possibility that formant amplitude patterns might be influenced by cultural stereotypes of speech styles. Perhaps a more interesting application of the theory of society's influence on the acquisition of gendered identities is apparent when we consider together the processes involved in separating the fundamental and formant frequencies of the sexes in adults and children. We know that in adults, both fundamental and formant frequencies are different across the sexes due to anatomical differences and, in the case of formants, there is also a component of voluntary manipulation of the relevant vocal organs. In children there are no significant fundamental frequency differences and we are left to speculate on the nature of the formant frequency differences which are measurable but small and inconsistent. Table 4.13 below will clarify this situation.

Table 4.13 The nature of perceptual differences between the sexes in adults and children

	Adults	Children
Fo	✓	✗
Formants	✓ + extra	?

Key

✓ = Acoustic difference caused by anatomical difference
✗ = No acoustic or anatomical difference
extra = Additional acoustic separation due to voluntary vocal tract manipulation

According to this descriptive model, we can restate two possibilities which we have already raised to complete the intersection of the 'children' and 'formants' sections of the table. The first possibility is that there is no size differences between the vocal tract lengths of prepubertal boys and girls and that the formant differences which we have encountered can be accounted for totally by means of sex-specific adaptations in speech style (i.e. cultural influences). In this case, there would be a simple mapping in table 4.13 of ✓ in adults (acoustic / anatomical difference) to ✗ in children (no difference). The lower right cell of the table would then be completed by '✗ + extra'. There is an elegance in this solution in that anatomical differences caused by sexual dimorphism are precluded until puberty and the influence of pressure to conform to sex roles appears to operate on formant frequencies in both adults and children.

The second possibility describes the situation in which formant frequency differences in the children are explained by positing differences prior to puberty in vocal tract lengths between the sexes. As we noted earlier there is an indication that boys may display slightly longer pharynges than girls during the years before puberty (King, 1952) however the evidence is far from conclusive and there appears to be no recorded differences in dimensions of the oral cavity at the same ages. Recalling also the discussion of vocal fold growth in section 1.2.1, the suggestion was made that a quantal rather than linear theory of growth appeared to fit the available data best. Whilst there is no *a priori* reason to assume that vocal tract growth patterns should precisely mirror vocal fold growth patterns, given the morphological proximity in the tissues involved there is probably some histological similarity in the growth schemes of the two systems. Furthermore, children have the opportunity to alter their Fo levels to synthesise a sex-difference where there is none yet they do not appear to do so. This could be regarded as an indication that they may choose, similarly, not to manipulate their formant frequencies (or to rephrase the issue, they may not have achieved the fine control over their acoustic output to make such a consistent alteration practical or possible).

In summary, there are two hypotheses which attempt to account for the differential acoustic separation of adults' and children's formant and fundamental frequencies. Both hypotheses hold that there are probably no anatomical differences in vocal fold size between the sexes in children and this accounts for the equality of Fo. One hypothesis suggests that the small sex-differences visible in children's formants are due to voluntary modifications in phonetic output and are therefore analogous to the situation in adults where, of course, there is also an anatomical difference. The other hypothesis suggests that, as children do not voluntarily alter their Fo towards perceptual separation, they may not alter their formant frequencies either and the differences must be due to a small, gradually widening anatomical gulf between boys and girls. Once again, without reliable and detailed measurements of the vocal tract lengths of prepubertal children we cannot make firm conclusions regarding the two possibilities although the 'voluntary adjustments' hypothesis appears to be the more likely of the two as the theory which proposes anatomical differences suffers from poor comparison with the vocal fold situation.

The most dramatic acoustic evidence to emerge from this work is the total equivalence between boys and girls in measures of fundamental frequency, pitch range, highest and lowest Fo, all perturbation measures, both short and long term and in the frequency and amplitude domains (jitter and shimmer). Clearly 4-5 year-old children are much more acoustically and physiologically similar than are adult males and females and yet the listener's perceptual capacity still manages to distinguish the sexes relatively successfully. This ability, although better than we might expect given the similarity of the vocal outputs of the boys and girls, is not perfect. Some children are not well identified by all judges and some judges perform more poorly than others. Clearly there is information being transferred from speaker to listener which depends on a number of factors including the development of the individual child in terms of physical maturity, linguistic progress and social sensitivity. At the same time, there are features of the listeners which contribute to the final identification scores including sex, receptive skills and personal experience. These two sources of variation, listener variability and speaker variability, interact to yield a continuum of results from very highly identified children to poorly identified children and from very accurate judges to less accurate judges.

Given our finding of a significant difference in correct identification scores between the boys and girls but not between male and female judges, we can conclude that the same judges would probably show the same pattern of results with different children. The semi-random sampling involved in the selection of children for the study allows us to extend the findings of this single experiment to similar samples of children drawn from the same population. We cannot, however, make a firm prediction of the result if different male and female judges were to judge the same children. As the variability amongst the judges was relatively high, there is no clear rationale upon which we can base a prediction, consequently it is difficult to draw conclusions regarding speakers and listeners in general. Nevertheless, we have revealed a wealth of evidence which shows that the gender of a large majority of children can be identified from voice alone at an age at which we must assume there is a minimum of sexual dimorphism. Furthermore, listeners vary as to their accuracy of gender identification with evidence of a sex-linked trend which favours females but does not firmly separate the group according to sex.

4.4 FUTURE DIRECTIONS

Whilst this study has undertaken to investigate and measure the ability of listeners to identify the gender of prepubertal children from samples of voice and has also shed light on the issue of gender differences in speaker characteristics, there are a number of areas of some relevance which have proven to be beyond the scope of the project. Furthermore, there are aspects of the methodology in particular which have been necessarily limited by the unavailability of certain techniques. This section will suggest some potential avenues of future research and highlight areas in which the study could be enhanced by means of different techniques.

One of the major problems encountered by the student or researcher in the field of vocal sex-differences is that there is a reasonable body of previous research, each item of which has been carried out to illuminate some aspect of the area. Unfortunately, these studies were often aimed at revealing slightly different features of the gender-bearing information and so were designed with different methodologies. Many of these studies show conflicting results (e.g. different studies to demonstrate the Fo characteristics of prepubertal children have shown girls to have higher, lower and equal mean Fo values to boys - respectively, Hasek, Singh and Murry, 1980; Sachs, Lieberman and Erickson, 1973; Bennett and Weinberg, 1979a).

Whilst this study has demonstrated the nature of the sex-identification ability across speakers and listeners for a sample of Scottish children, we cannot make firm predictions about the process in general terms unless we can repeat our experiments with children of different ages and, ideally, from different cultural backgrounds. Some of the variability in the data yielded by previous research has originated from the fact that many different researchers used children with very different ages, assuming that all pre-pubertal children would show identical results. This has turned out to be unfortunate as we believe that the child prior to puberty is developing in a dynamic and possibly sex-differential manner (i.e. boys and girls do not mature at identical rates). The speech data from a 3½ year old will differ in important ways from the speech of a 9 year old and yet both are thought of as simply prepubertal.

Clearly what is needed is a large-scale study which performs exactly the same analyses on exactly the same type of speech samples taken from groups of children with overlapping age ranges. Using this method it will become apparent exactly how young children's speech matures and to what extent this development is subject to sex-linked influences.

It was stated above that children of different cultures should be included in any large-scale replication of this study. This brings us on to an issue which has been discussed earlier in this chapter but again without tackling it with a specific methodology it cannot be settled here. The issue is of whether the sex-differences revealed in the present study and in its predecessors have their origin in biological or cultural influences. The discussion of the present study has hinted that the balance of responsibility probably lies with environmental or cultural influences rather than innate ones however this is, at best, informed speculation. A balanced sociolinguistic study which involved a large number of prepubertal boys and girls from a carefully selected sample of cultures from around the world would go a long way towards supplying some of the missing answers. Different cultures have developed different patterns of social behaviour and this includes child-rearing practices. When we consider valued male and female traits, we can see firstly that there are far more typically male traits which are valued by society than typically female traits. This contributes to the power mismatch between the sexes and enhances the male's position as dominant with respect to the female. In British culture, for example, positive traits which are fostered (rightly or wrongly) in the young male include being active, aggressive, competitive, independent, logical, objective and self-confident, whilst at the same time he is discouraged from showing his emotions too easily, crying in public, getting overexcited about a minor issue and being conceited about his appearance. Similarly there are a few traits which are valued in females such as empathy towards other peoples' feelings, gentleness, concern for one's own physical appearance, quietness, talkativeness and general neatness. The qualities which are thought of as being particularly unsuitable for females tend to correspond to the ones which are most highly valued amongst males. Whilst we can see that there are stereotypes of characteristic male and female attributes, we should remember that the average male or female is not a collection of features from one or other side of a list but rather an amalgam of typical male and female preferential traits with the balance normally biased towards those of the relevant gender. The communication patterns of males and females are undoubtedly shaped by the different personality qualities attributed to each by society and the construction of the male and female psyche are partially dependent on (and partly feed into) the expectations and belief systems which have been set up over centuries of social and cultural evolution.

These characteristics are based on observation taken from children in Western cultures and therefore are not necessarily universal (Maltz and Borker, 1982; Coates, 1995). A comparative study which investigates the interactions between parent and child and also the roles played by the sexes in adulthood throughout different cultures would be invaluable to our understanding of the cultural influence on sex-differences in communication.

On a more detailed level, there is scope for a deeper investigation into the phenomenon of breathiness. The two parameters assessed in this study (H1-F1 and H1-H2) are accurate indications of vocal breathiness in terms of spectral slope measures however the involvement of high-frequency noise components and glottal leakage has been overlooked. Furthermore, due to the influence of the vocal tract transfer function (formants) it has only been possible to consider data recovered from the vowel [a]. An experiment into breathiness which can make use of the technique of inverse filtering of the vocal spectrum will benefit from the ability to use any vowel as stimulus. It is therefore suggested that if future research considers that the area of phonation type, especially breathiness, is of continuing interest there should be attention paid to both spectral and airflow aspects of the phenomenon and digital processing techniques including inverse filtering should be used to prepare the speech samples. The technique of laryngography might reveal differences in phonatory behaviour between prepubertal boys and girls, however Robb and Simmons (1990) found minimal sex-differences in measures of vocal fold contact using electroglottography.

The notion of extending the research to include other parameters of interest extends to the acoustic experimentation carried out in this study. Whilst a total of 27 acoustic parameters[31] was extracted and analysed from the various speech samples, these only represented those parameters which were considered to have a good chance of representing a measurable gender difference. Clearly there are many more parameters which are open to investigation and a number of the same parameters which might benefit from a more intense study.

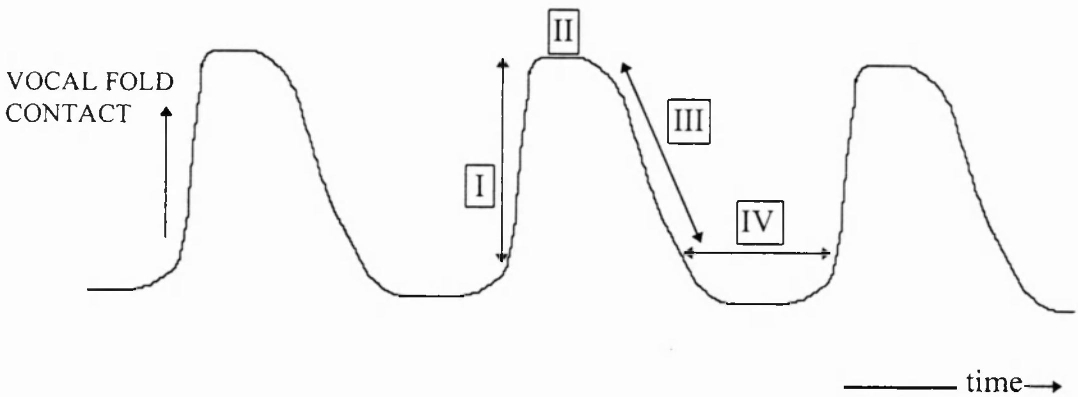
Some measure of vocal attack, if it could be implemented, would perhaps be revealing in the search for a gender-distinguishing parameter. Subjective impressions of the recordings of the children in the present study included comments relating to the 'strength' or 'masculinity' of some of the boys' voices. It has always proven difficult to extract reliable data concerning amplitude and power levels of children's voices. Böhme and Stuchlik (1995: 307) state that "voice profiles cannot be established for children below the age of 7 years because of the concentration and perceptive faculty necessary". Whilst it proved impossible to measure raw vocal amplitude from the recordings in this experiment, a methodology which was orientated more specifically and single-mindedly towards output level differences could possibly capture such information. The jitter and shimmer measurements in the present study failed to reveal any gender differences however there may yet be differences in perceptual roughness between the sexes which require a more sensitive gauge of quality of tone than is afforded by either jitter or shimmer measurements.

[31] These comprise the 18 parameters extracted from the sentence sample (see appendix 3) and frequencies, bandwidths and amplitudes of the first, second and third formants

Finally, there has been very little attention paid in previous studies to gender differences in communication of the differing abilities of male and female listeners. In this study, although not statistically significant in the final analysis, an indication that females may be better than males at identifying prepubertal children's gender from voice was forthcoming. The effect discovered should be large enough to provoke interest in investigating the precise nature of the gender identification ability. The question still remains to a large extent, is there really a sex-difference between men and women when it comes to recognition of gender. Although the statistics suggest that we should not interpret the findings as evidence of a real effect, it is difficult to dismiss the fact that the female judges consistently scored more correct judgements of gender than male judges across both boys and girls and in all speech sample types. For the moment the conclusions we can draw from the listener sex-difference results must be tentative, however a larger battery of tests which expose more of the truth surrounding the perceptual abilities of normal men and women must surely be overdue.

Appendix 1

Figure A1.1 An idealised laryngograph output



I closing phase
II maximum contact
III opening phase
} CLOSED PHASE

IV open phase — OPEN PHASE

The open and closed quotient of the waveform is the duration of the open or closed phase respectively divided by the total length of the cycle (Tx) and expressed as a percentage.

Open quotient = ((Open Phase / Tx) x 100)%

Closed quotient = ((Closed Phase / Tx) x 100)%

Appendix 2A

Queen Margaret College
EDINBURGH established 1875

Direct Dial - 317 3687

22 March, 1993

Dear Parent,

I am a research student, based at Queen Margaret College, conducting a study into aspects of voice quality in normal children. In order to carry out this work I hope to record samples of speech from a number of normally developing children who are aged between $4\frac{1}{2}$ and $5\frac{1}{2}$ (I regret that children who are being seen by a speech and language therapist cannot be included in the study).

The staff at the school have kindly agreed to co-operate in the study and the project has the full backing of the Lothian Regional Council Department of Education. I intend to spend some time in the classroom with the children prior to the recordings to allow them to become accustomed to my presence (and, of course, me to them!) The tape recording itself will last only around fifteen minutes and will consist of a number of simple word and story games.
The recorded speech will provide information regarding the speech of normally developing children which will be of use to speech therapists in the treatment of speech disorders.
If you would be willing for your child to participate in this study would you please complete the attached consent form and return it to the head-teacher, Mrs. Jack.

In order to avoid taking up the teachers' valuable time, I would be grateful if you would address any questions or matters you might like to discuss concerning the project to myself and not to the school staff. I am available during office hours at the above telephone number and my supervisors and I are more than happy to answer any questions.

Yours sincerely,

Moray Nairn	Dr. Janet Beck	Dr. Nigel Hewlett
Researcher	Lecturer in Speech Pathology and Therapy	Lecturer in Linguistics
	(Supervisor)	**(Supervisor)**

AGREEMENT TO TAKE PART IN VOICE QUALITY STUDY

I, (name) ....
Of (address) ....
....

fully and freely give consent for my son/daughter .... to take part in the study named above. I understand that the study is designed to try to gain information about aspects of normal voice quality. I also realise that the identity of my child shall remain undisclosed and that the confidentiality of any material relating to him/her shall be maintained.
I further agree that the recordings made in the course of this study may be used in any presentations resulting from this work.
I note that I may freely decide at any stage to withdraw my child from the study.

Signed : ....
Date : ....

Date of Birth of Child : ....

Appendix 2

CHILDREN'S VOICE AND GENDER EXPERIMENT

General Instructions

This experiment is designed to yield information regarding the features in the voice that allow us to identify the gender of a speaker. The form of the experiment is a series of listening tests in which all of the speakers are children between the ages of 4½ and 5½ years and the listeners are young male and female adults.

As a listener it is your task to listen to each of the pieces of recorded speech and decide for each whether you think that the speaker is a boy or a girl. There are three separate tapes to listen to. The first tape consists of 89 children each speaking three isolated vowel sounds, therefore once you have heard each child speak his or her three vowels you must mark on the response sheet whether you think that the speaker is a boy or a girl. The second tape consists of 89 children each speaking the same two sentences, again once you have heard the two sentences for each child you must decide whether it is a boy or a girl speaking. The third tape consists of 89 children speaking short passages of connected speech. Each passage will be slightly different but the topic being discussed is the same in every case. Once you have listened to the child's spoken passage you must mark your judgement of his or her gender on the response sheet.

Tape 1 (vowels) will last 50 minutes
Tape 2 (sentences) will last 35 minutes
Tape 3 (passages) will last 50 minutes

There are further instructions at the top of each response sheet. Before starting this experiment all listeners **must** sign a consent form.
The fee for completion of the experiment (i.e. listening and responding to all 3 tapes) is £10.

If you have any questions please ask them **NOW**.

CHILDREN'S VOICE AND GENDER EXPERIMENT

Tape 1 - vowels

You are about to hear 89 children from Edinburgh primary schools (all aged between 4½ and 5½ years). Each of them will speak three vowels - the first vowel is **ee**, like the vowel in the word *bee*; the second vowel is **a**, like the vowel in the word *hat*; and the third vowel is **oh**, like the vowel in the word *go*.
You will hear the vowels in order with a 5 second gap between each one. There is a 15 second gap between each child during which you should mark your judgement of the child's gender by circling M for male or F for female. Please ensure that you circle one and only one for each child. If you are in doubt as to whether the child is male or female then make a guess at it. **Do not leave it blank**.
There is a one minute gap after child number 30 and after child number 60 during which you may rest and also to help you keep track of your position on the tape.

1	M	F
2	M	F
3	M	F
4	M	F
5	M	F
6	M	F
7	M	F
8	M	F
9	M	F
10	M	F
11	M	F
12	M	F
13	M	F
14	M	F
15	M	F
16	M	F
17	M	F
18	M	F
19	M	F
20	M	F
21	M	F
22	M	F
23	M	F
24	M	F
25	M	F
26	M	F
27	M	F
28	M	F
29	M	F
30	M	F

31	M	F
32	M	F
33	M	F
34	M	F
35	M	F
36	M	F
37	M	F
38	M	F
39	M	F
40	M	F
41	M	F
42	M	F
43	M	F
44	M	F
45	M	F
46	M	F
47	M	F
48	M	F
49	M	F
50	M	F
51	M	F
52	M	F
53	M	F
54	M	F
55	M	F
56	M	F
57	M	F
58	M	F
59	M	F
60	M	F

61	M	F
62	M	F
63	M	F
64	M	F
65	M	F
66	M	F
67	M	F
68	M	F
69	M	F
70	M	F
71	M	F
72	M	F
73	M	F
74	M	F
75	M	F
76	M	F
77	M	F
78	M	F
79	M	F
80	M	F
81	M	F
82	M	F
83	M	F
84	M	F
85	M	F
86	M	F
87	M	F
88	M	F
89	M	F

CHILDREN'S VOICE AND GENDER EXPERIMENT

Tape 2 - sentences

You are about to hear 89 children from Edinburgh primary schools (all aged between 4½ and 5½ years). Each of them will speak two sentences - the first sentence is **"Rover is a big brown dog"** and the second sentence is "**Rover has got a bone in his mouth**". The children occasionally say a wrong word when trying to produce the correct sentences but this is not important.

You will hear the sentences in order with a 5 second gap between the first and the second one. There is a 10 second gap between each child during which you should mark your judgement of the child's gender by circling M for male or F for female. Please ensure that you circle one and only one for each child. If you are in doubt as to whether the child is male or female then make a guess at it. **Do not leave it blank**. There is a one minute gap after child number 30 and after child number 60 during which you may rest and also to help you keep track of your position on the tape.

1	M	F
2	M	F
3	M	F
4	M	F
5	M	F
6	M	F
7	M	F
8	M	F
9	M	F
10	M	F
11	M	F
12	M	F
13	M	F
14	M	F
15	M	F
16	M	F
17	M	F
18	M	F
19	M	F
20	M	F
21	M	F
22	M	F
23	M	F
24	M	F
25	M	F
26	M	F
27	M	F
28	M	F
29	M	F
30	M	F

31	M	F
32	M	F
33	M	F
34	M	F
35	M	F
36	M	F
37	M	F
38	M	F
39	M	F
40	M	F
41	M	F
42	M	F
43	M	F
44	M	F
45	M	F
46	M	F
47	M	F
48	M	F
49	M	F
50	M	F
51	M	F
52	M	F
53	M	F
54	M	F
55	M	F
56	M	F
57	M	F
58	M	F
59	M	F
60	M	F

61	M	F
62	M	F
63	M	F
64	M	F
65	M	F
66	M	F
67	M	F
68	M	F
69	M	F
70	M	F
71	M	F
72	M	F
73	M	F
74	M	F
75	M	F
76	M	F
77	M	F
78	M	F
79	M	F
80	M	F
81	M	F
82	M	F
83	M	F
84	M	F
85	M	F
86	M	F
87	M	F
88	M	F
89	M	F

Tape 3 - passages

You are about to hear 89 children from Edinburgh primary schools (all aged between 4½ and 5½ years). Each of them will speak an extract from a story which they had been told earlier, therefore each child's telling of the story is slightly different but the topic is the same for each.

You will hear approximately 20 seconds of the story spoken by each child. It has occasionally been necessary to splice together some short phrases spoken by an individual child because that child may not have spoken 20 seconds of fluent connected speech. There is a 10 second gap between each child during which you should mark your judgement of the child's gender by circling M for male or F for female. Please ensure that you circle one and only one for each child. If you are in doubt as to whether the child is male or female then make a guess at it. **Do not leave it blank**.

There is a one minute gap after child number 30 and after child number 60 during which you may rest and also to help you keep track of your position on the tape.

1	M	F
2	M	F
3	M	F
4	M	F
5	M	F
6	M	F
7	M	F
8	M	F
9	M	F
10	M	F
11	M	F
12	M	F
13	M	F
14	M	F
15	M	F
16	M	F
17	M	F
18	M	F
19	M	F
20	M	F
21	M	F
22	M	F
23	M	F
24	M	F
25	M	F
26	M	F
27	M	F
28	M	F
29	M	F
30	M	F

31	M	F
32	M	F
33	M	F
34	M	F
35	M	F
36	M	F
37	M	F
38	M	F
39	M	F
40	M	F
41	M	F
42	M	F
43	M	F
44	M	F
45	M	F
46	M	F
47	M	F
48	M	F
49	M	F
50	M	F
51	M	F
52	M	F
53	M	F
54	M	F
55	M	F
56	M	F
57	M	F
58	M	F
59	M	F
60	M	F

61	M	F
62	M	F
63	M	F
64	M	F
65	M	F
66	M	F
67	M	F
68	M	F
69	M	F
70	M	F
71	M	F
72	M	F
73	M	F
74	M	F
75	M	F
76	M	F
77	M	F
78	M	F
79	M	F
80	M	F
81	M	F
82	M	F
83	M	F
84	M	F
85	M	F
86	M	F
87	M	F
88	M	F
89	M	F

Thank you for participating

Appendix 3

The acoustic parameters extracted from the sentence samples are listed below. Full definitions of each parameter are accompanied by details of the algorithm(s) used to extract each. Voice break areas were automatically excluded from the calculation of all parameters.

- Fundamental Frequency Information (Average Fo, Average Pitch Period, Highest Fo, Lowest Fo, Standard deviation of Fo, Phonatory Fo Range, Number of Pitch Periods computed)
- Short and Long Term Frequency Perturbation Information (Absolute Jitter, Jitter Percent, Relative Average Perturbation, Pitch Period Perturbation Quotient, Smoothed PPQ, Fundamental Frequency Variation)
- Short and Long Term Amplitude Perturbation Information (Shimmer in dB, Shimmer Percent, Amplitude Perturbation Quotient, Smoothed APQ, Peak Amplitude Variation)

Fundamental Frequency Information

The Multi-Dimensional Voice Processor (MDVP) uses an adaptive time-domain pitch synchronous method for pitch extraction. This involves the following steps:

1. The voice signal is windowed (divided into equal sized time chunks) with window size set at 30 msec.
2. A pitch-adaptive modified autocorrelation function is computed for every window of voice data. The pitch period for every window is extracted and voiced / unvoiced decisions are made.
3. A decision-making routine is used to correct the errors in the extracted pitch caused by sub-harmonics, severe irregularities of the pitch period, noise, voice breaks and improper voiced / unvoiced decisions.

4. An accurate period-to-pitch detection is made using peak-to-peak extraction. It is synchronous with the corrected pitch and voiced / unvoiced results computed earlier.

5. A linear 5-point interpolation is applied on the final period-to-period pitch data in order to increase the resolution. This is essential for the jitter measurements.

Fo

Definition: Average Fundamental Frequency (Hz). This parameter represents the average value of all of the extracted period-to-period fundamental frequency values.

Method: Fo is computed from the extracted period-to-period pitch data as:

$$Fo = \frac{1}{N}\sum_{i=1}^{N} Fo^{(i)}$$

where: $Fo^{(i)} = \dfrac{1}{To^{(i)}}$ (period-to-period fundamental frequency)

$To^{(i)}, i = 1,2 \ldots N$ (extracted pitch period data)

N = Number of extracted pitch periods

To

Definition: Average Pitch Period (msec.)

Method: To is computed as:

$$To = \frac{1}{N}\sum_{i=1}^{N} To^{(i)}$$

where: $To^{(i)}, i = 1,2 \ldots N$ (extracted pitch period data)

N = Number of extracted pitch periods

Fhi

Definition: Highest Fundamental Frequency (Hz). This parameter represents the greatest of all period-to-period fundamental frequency values.

Method: *Fhi* is computed as:

$$Fhi = \max\left\{Fo^{(i)}\right\}, i = 1,2\ldots N$$

where: $Fo^{(i)} = \dfrac{1}{To^{(i)}}$ (period-to-period fundamental frequency)

$To^{(i)}, i = 1,2\ldots N$ (extracted pitch period data)

N = Number of extracted pitch periods

Notes: The pitch extraction range is limited to 70-625 Hz, therefore the system will not identify *Fhi* > 625 Hz.

Flo

Definition: Lowest Fundamental Frequency (Hz). This parameter represents the smallest of all period-to-period fundamental frequency values.

Method: *Flo* is computed as:

$$Flo = \min\left\{Fo^{(i)}\right\}, i = 1,2\ldots N$$

where: $Fo^{(i)} = \dfrac{1}{To^{(i)}}$ (period-to-period fundamental frequency)

$To^{(i)}, i = 1,2\ldots N$ (extracted pitch period data)

N = Number of extracted pitch periods

Notes: The pitch extraction range is limited to 70-625 Hz, therefore the system will not identify *Flo* < 70 Hz.

STD

Definition: Standard deviation of all extracted period-to-period fundamental frequency values (Hz).

Method: *STD* is computed as standard deviation of the extracted period-to-period fundamental frequency data as:

$$STD = \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Fo - Fo^{(i)})^2}$$

where: $$Fo = \frac{1}{N}\sum_{i=1}^{N} Fo^{(i)}$$

N = number of extracted pitch periods

PFR

Definition: Phonatory Fundamental Frequency Range (Semi-tones). This parameter represents the range between *Fhi* and *Flo* expressed in number of semi-tones.

Method: *PFR* is computed by accounting the number of semi-tones in the frequency range *Flo* to *Fhi*. The ration of two consecutive semi-tones is equal to 12th root of 2.

All frequencies of semi-tones $Fst^{(k)} = f_1 a^k, k = 1,2...$ are computed within the frequency range 55 Hz to 1055 Hz.

where: $a = \sqrt[12]{2}$,

$$f_1 = 55Hz, f_2 = 1055Hz \text{ and } f_1 \le Fst^{(k)} \le f_2$$

Short and Long Term Frequency Perturbation Information

Jita

Definition: Absolute Jitter (μsec). This parameter represents an evaluation of the period-to-period variability of the pitch period within the analyzed voice sample.

Method: *Jita* is computed as:

$$Jita = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|To^{(i)} - To^{(i+1)}\right|$$

where: $To^{(i)}, i = 1,2 \ldots N$ (extracted pitch period data)

N = number of extracted pitch periods

Notes: Absolute jitter measures the very short-term (cycle-to-cycle) irregularity of the pitch periods in the voice. This measure is widely used in the research literature on voice perturbation. It is very sensitive to the pitch variations occurring between consecutive pitch periods.
Both *Jita* and *Jitt* represent evaluations of the same type of pitch perturbation. *Jita* is an absolute measure and shows the result in microseconds which makes it dependent on the fundamental frequency of the voice. For this reason, the normative values of *Jita* for adult men and women differ significantly. Higher pitch tends to result in lower *Jita*. *Jita* values of two subjects with different speaking Fo levels are therefore difficult to compare. *Jitt* is a relative measure and the influence of the average Fo of the subject is significantly reduced.

Jitt

Definition: Jitter Percent (%). This parameter represents the relative evaluation of the period-to-period variability of the pitch period within the analyzed voice sample.

Method: *Jitt* is computed as:

$$Jitt = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|To^{(i)} - To^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N}To^{(ii)}}$$

where: $To^{(i)}, i = 1,2 \ldots N$ (extracted pitch period data)

N = number of extracted pitch periods

Notes: Jitter Percent measures the very short-term (cycle-to-cycle) irregularity of the pitch periods in the voice. This measure is widely used in the research literature on voice perturbation. It is very sensitive to the pitch variations occurring between consecutive pitch periods.
Both *Jita* and *Jitt* represent evaluations of the same type of pitch perturbation. *Jita* is an absolute measure and shows the result in microseconds which makes it dependent on the fundamental frequency of the voice. For this reason, the normative values of *Jita* for adult men and women differ significantly. Higher pitch tends to result in lower *Jita*. *Jita* values of two subjects with different speaking Fo levels are therefore difficult to compare. *Jitt* is a relative measure and the influence of the average Fo of the subject is significantly reduced.

RAP

Definition: Relative Average Perturbation (%). This parameter represents the relative evaluation of the period-to-period variability of the pitch with smoothing factor of 3 periods.

Method: *RAP* is computed as:

$$RAP = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1}\left|\frac{To^{(i-1)}+To^{(i)}+To^{(i+1)}}{3} - To^{(i)}\right|}{\frac{1}{N}\sum_{i=1}^{N} To^{(i)}}$$

where: $To^{(i)}, i = 1,2 \ldots N$ (extracted pitch period data)

N = Number of extracted pitch periods

Notes: Relative Average Perturbation measures the short term (cycle-to-cycle with smoothing factor of 3 periods) irregularity of the pitch period of the voice. The smoothing reduces the sensitivity of *RAP* to pitch extraction errors. Although sensitive to short term perturbation, it is less sensitive than *Jitt* or *Jita* to very short term period-to-period variations. An increased value of *RAP* may be associated with hoarse or breathy voice.

---

sPPQ

Definition: Smoothed Pitch Period Perturbation Quotient (%). This parameter represents the relative evaluation of the short or long-term variability of the pitch period within the voice sample at a smoothing factor chosen by the researcher.

Method: *sPPQ* is computed as:

$$sPPQ = \frac{\frac{1}{N-sf+1}\sum_{i=1}^{N-sf+1}\left|\frac{1}{sf}\sum_{r=0}^{sf-1} To^{(i+r)} - To^{(i+m)}\right|}{\frac{1}{N}\sum_{i=1}^{N} To^{(i)}}$$

where: $To^{(i)}, i = 1,2 \ldots N$ (extracted pitch period data)

N = Number of extracted pitch periods
sf = smoothing factor chosen by user

Notes:

sPPQ allows the user to define a pitch perturbation measure by selecting the smoothing factor from 1 to 199.
With a small smoothing factor, *sPPQ* is sensitive mostly to the short-term pitch variations of the voice. A smoothing factor of 1 (i.e. no smoothing) equates *sPPQ* with *Jitt* (Jitter Percent). A smoothing factor of 3 equates *sPPQ* with *RAP* (Relative Average Pitch Perturbation). *sPPQ* with a smoothing factor of 5 represents the parameter described as *PPQ*. Because of the smoothing, *PPQ* and *RAP* are less sensitive to the period-to-period pitch variations, however they describe short-term jitter well.
At high smoothing factors *sPPQ* correlates with the intensity of the long-term pitch period variations. (see *vFo* notes)

---

vFo

Definition:

Coefficient of Fundamental Frequency Variation (%).
This parameter represents the relative standard deviation of the fundamental frequency. It reflects, in general, the variation of Fo (short-to-long term) within the voice sample.

Method:

vFo is computed as the ratio of the standard deviation of the extracted period-to-period fundamental frequency data to the average fundamental frequency as:

$$vFo = \frac{\sigma}{Fo} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{N}\sum_{j=1}^{N} Fo^{(j)} - Fo^{(i)})^2}}{\frac{1}{N}\sum_{i=1}^{N} Fo^{(i)}}$$

where: $Fo = \frac{1}{N}\sum_{i=1}^{N} Fo^{(i)}$

and $Fo^{(i)} = \frac{1}{To^{(i)}}$ (period-to-period fundamental frequency)

and $To^{(i)}, i = 1,2 \ldots N$ (extracted pitch period data)

and N = Number of extracted pitch periods

Short and Long Term Amplitude Perturbation Information

Shimmer measures are widely used in the research literature on voice perturbation. Both *Shim* and *ShdB* are sensitive to the amplitude variations occurring between consecutive pitch periods, however, pitch extraction errors may affect both parameters significantly. *APQ* is often preferred as a measure for shimmer because it is less sensitive to pitch extraction errors while still providing a reliable indication of the short-term amplitude variability in the voice.

ShdB

Definition: Shimmer in decibels (dB). This parameter represents the evaluation of the period-to-period (very short-term) variability of the peak-to-peak amplitude within the voice.

Method: *ShdB* is computed as:

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20\log\left(A^{(i+1)} / A^{(i)}\right)\right|$$

where: $A^{(i)}, i = 1,2 \ldots N$ (extracted peak-to-peak amplitude data)

N = number of extracted impulses

Shim

Definition: Shimmer Percent (%). This parameter represents the relative evaluation of the period-to-period (very short-term) variability of the peak-to-peak amplitude within the voice.

Method: *Shim* is computed as:

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A^{(i)} - A^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}}$$

where: $A^{(i)}, i = 1,2 \ldots N$ (extracted peak-to-peak amplitude data)

N = number of extracted impulses

sAPQ

Definition: Smoothed Amplitude Perturbation Quotient (%). This parameter represents the evaluation of the long or short-term variability of the peak-to-peak amplitude in the voice at a smoothing factor chosen by the researcher.

Method: *sAPQ* is computed as:

$$sAPQ = \frac{\frac{1}{N-sf+1}\sum_{i=1}^{N-sf+1}\left|\frac{1}{sf}\sum_{r=0}^{sf-1}A^{(i+r)} - A^{(i+m)}\right|}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}}$$

where: $A^{(i)}, i = 1,2...N$ (extracted peak-to-peak amplitude data)

N = number of extracted impulses

Notes: *sAPQ* allows the user to define an amplitude perturbation measure by selecting the smoothing factor from 1 to 199. With a small smoothing factor, *sAPQ* is sensitive mostly to the short-term amplitude variations of the voice. A smoothing factor of 1 (i.e. no smoothing) equates *sAPQ* with *Shim* (Shimmer Percent). *sAPQ* with a smoothing factor of 11 represents the parameter described as *APQ*. Because of the smoothing, *APQ* is less sensitive to the period-to-period pitch variations, however it still describes short-term shimmer well. At high smoothing factors *sAPQ* correlates with the intensity of the long-term peak-to-peak amplitude variations.

vAM

Definition: Coefficient of Amplitude Variation (%). This parameter represents the relative standard deviation of the peak-to-peak amplitude. It reflects in general the amplitude variations (short to long-term) within the voice.

Method: *vAm* is computed as the ratio of the standard deviation to the average value of the extracted peak-to-peak amplitude data as:

$$vAm = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N}A^{(j)} - A^{(i)}\right)^2}}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}}$$

where: $A^{(i)}, i = 1,2\ldots N$ (extracted peak-to-peak amplitude data)

N = number of extracted impulses

Appendix 4

The results of the four Scheffé tests carried out on the material type variable are shown below.

1) All judges - By children (see p. 85)

	Vowels	Passage	Sentences
Vowels	0.000000	5.999931	9.187469
Passage	0.000000	0.000000	3.187538
Sentences	0.000000	0.000000	0.000000

Observations per cell = 89
Error Degrees of Freedom = 174
Mean Square Error = 0.133000
$F(2,174)$ for $P < 0.05$ = 3.030000
$F(2,174)$ for $P < 0.01$ = 4.700000

Critical value for $P < 0.05$ = 11.977664
Critical value for $P < 0.01$ = 14.917627

2) All judges - By Judges (see p. 88)

	Vowels	Passage	Sentences
Vowels	0.000000	1.073760	1.651600
Passage	0.000000	0.000000	0.577840
Sentences	0.000000	0.000000	0.000000

Observations per cell = 16
Error Degrees of Freedom = 28
Mean Square Error = 0.008000
$F(2,28)$ for $P < 0.05$ = 3.340000
$F(2,28)$ for $P < 0.01$ = 5.450000

Critical value for $P < 0.05$ = 1.307700
Critical value for $P < 0.01$ = 1.670449

3) Judge f1 removed - By children (see p. 102)

	Vowels	Passage	Sentences
Vowels	0.000000	6.893051	9.087349
Passage	0.000000	0.000000	2.194298
Sentences	0.000000	0.000000	0.000000

Observations per cell = 89
Error Degrees of Freedom = 174
Mean Square Error = 0.137000
$F(2,174)$ for $P < 0.05$ = 3.030000
$F(2,174)$ for $P < 0.01$ = 4.700000

Critical value for P < 0.05 = 12.156445
Critical value for P < 0.01 = 15.140290

4) Judge f1 removed - By Judges (see p. 103)

	Vowels	Passage	Sentences
Vowels	0.000000	1.167450	1.552950
Passage	0.000000	0.000000	0.385500
Sentences	0.000000	0.000000	0.000000

Observations per cell = 15
Error Degrees of Freedom = 26
Mean Square Error = 0.007000
$F(2,26)$ for $P < 0.05$ = 3.370000
$F(2,26)$ for $P < 0.01$ = 5.530000

Critical value for P < 0.05 = 1.189706
Critical value for P < 0.01 = 1.524008

Appendix 5

Figure A5.1 Breakdown of judgements of individual female children

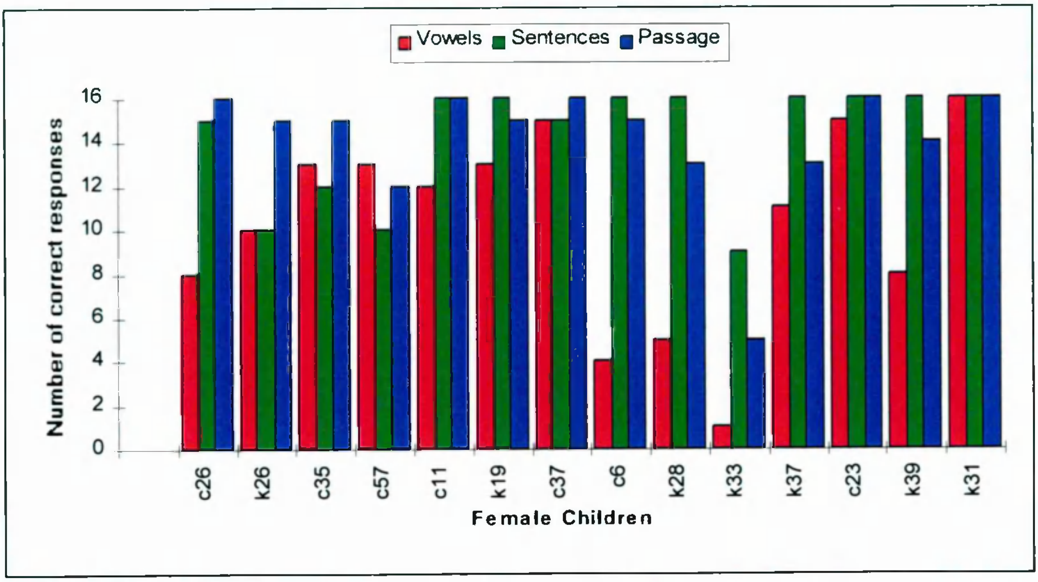


Figure A5.2 Breakdown of judgements of individual female children (contd)

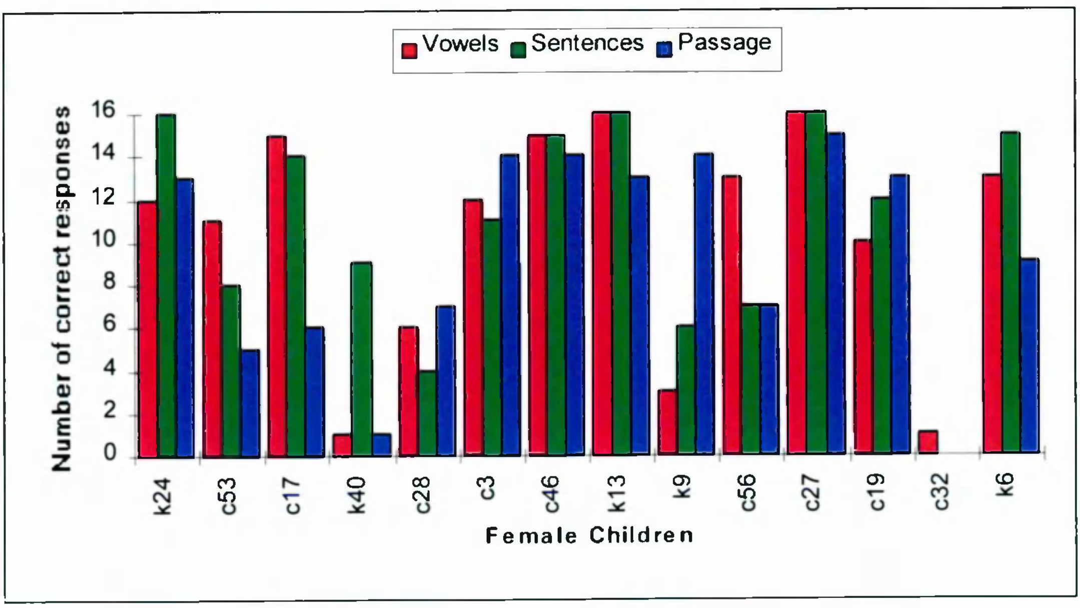


Figure A5.3 Breakdown of judgements of individual female children (contd)

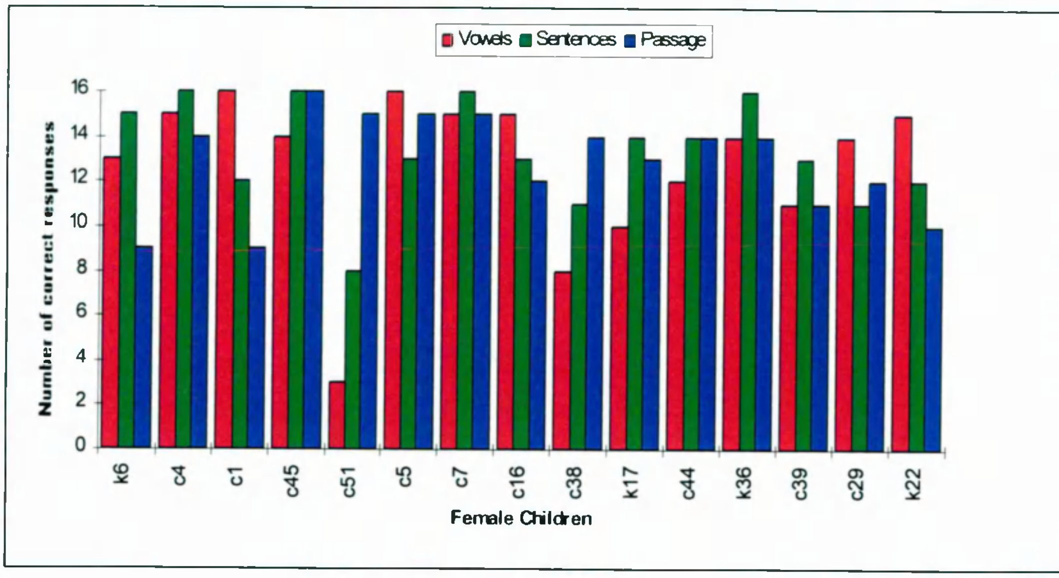


Figure A5.4 Breakdown of judgements of individual male children

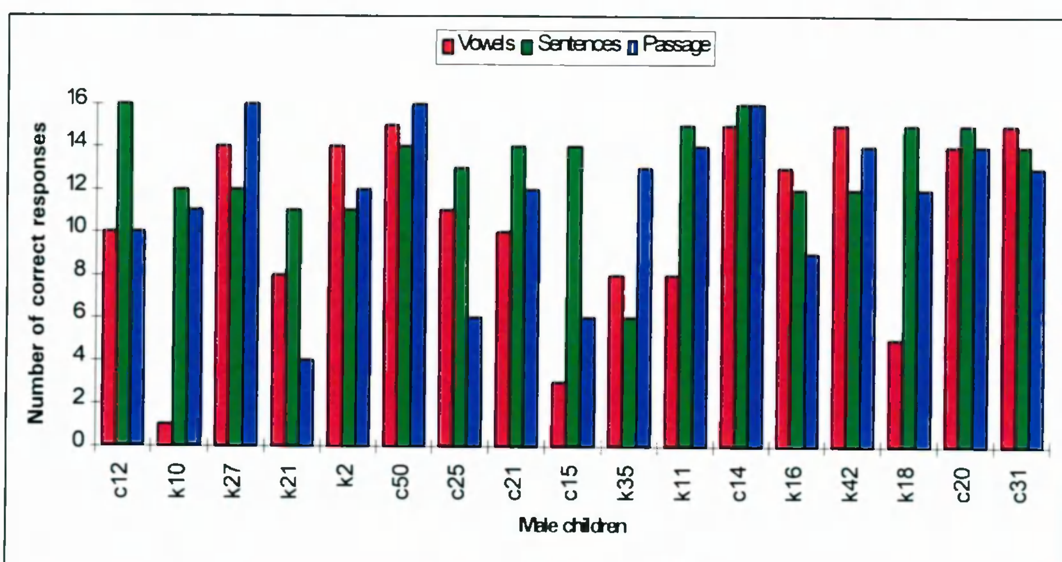


Figure A5.5 Breakdown of judgements of individual male children (contd)

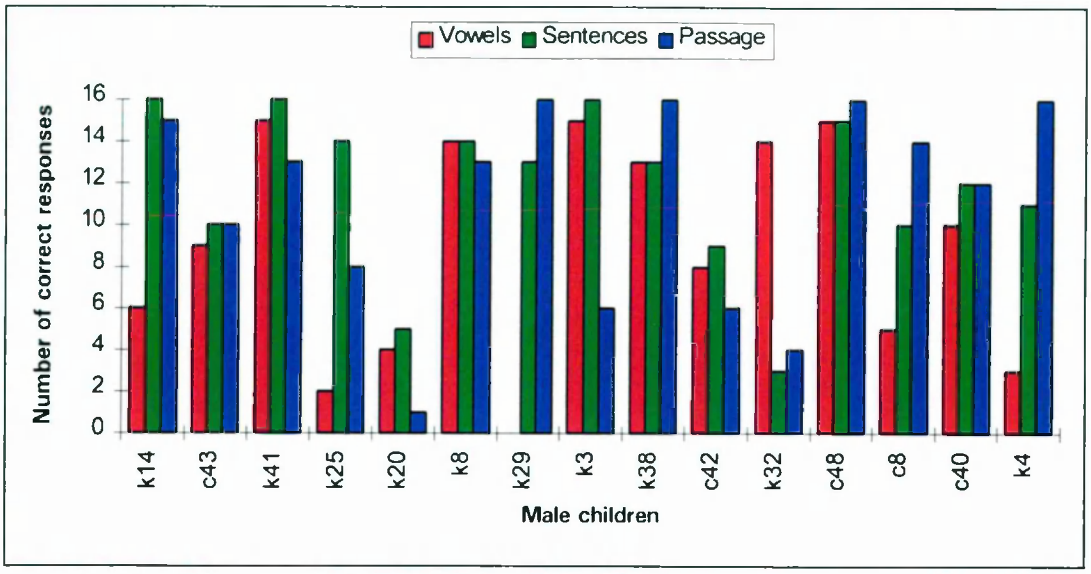
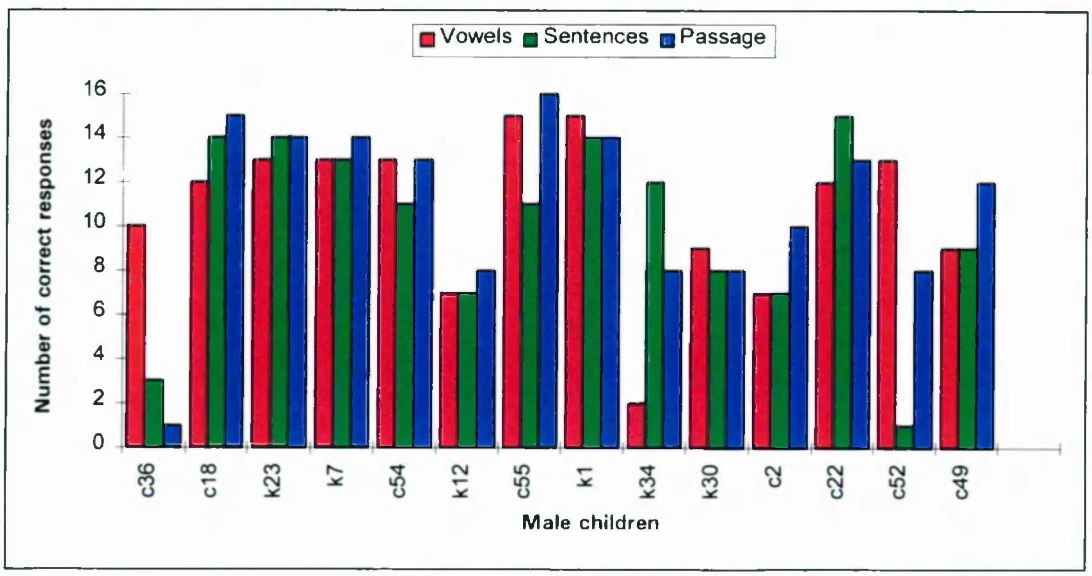


Figure A5.6 Breakdown of judgements of individual male children (contd)



Appendix 6

Table A6.1 Results of a two-factor Analysis of Variance between values of log β of male and female judges for all sample types with the data from judge f1 removed.

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Samples	0.037653	2	0.018826	2.160792	0.129948	3.259444
Sex of Judge	1.76E-05	1	1.76E-05	0.00202	0.964403	4.113161
Interaction	0.001156	2	0.000578	0.066361	0.935907	3.259444
Within	0.313656	36	0.008713			
Total	0.352483	41				

Table A6.2 Results of a two-factor Analysis of Variance between values of d' of male and female judges for all sample types with the data from judge f1 removed.

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Samples	3.42621	2	1.713105	12.64315	6.93E-05	3.259444
Sex of Judge	0.143618	1	0.143618	1.059934	0.310098	4.113161
Interaction	0.139658	2	0.069829	0.515357	0.601627	3.259444
Within	4.877882	36	0.135497			
Total	8.587369	41				

Appendix 7

Charts showing mean formant differences between the most male-like and the least male-like children.

Figure A7.1 First formant information of [a]in most and least male-like children

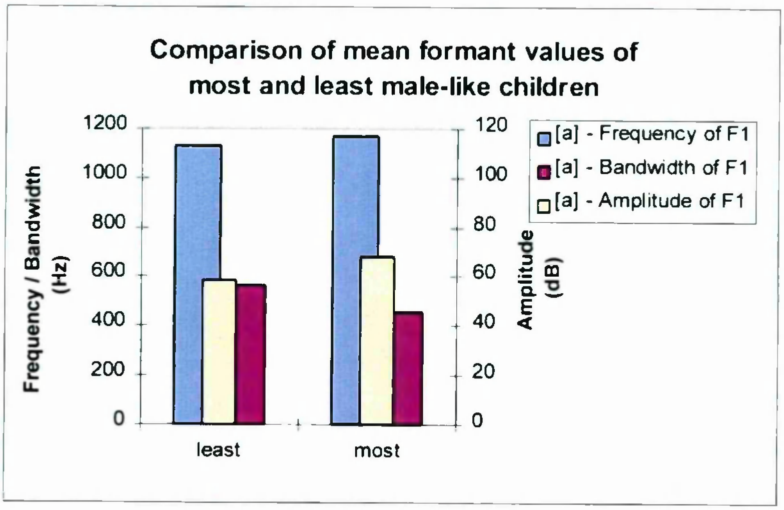


Figure A7.2 Second formant information of [a] in most and least male-like children

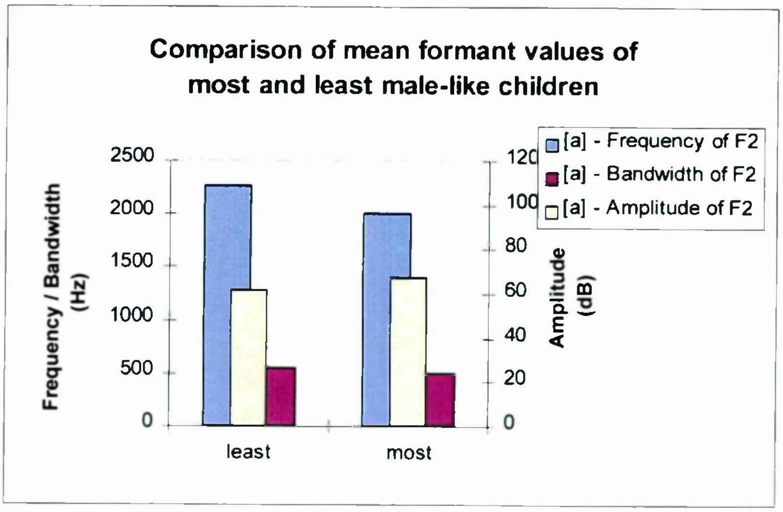


Figure A7.3 Third formant information of [a] in most and least male-like children

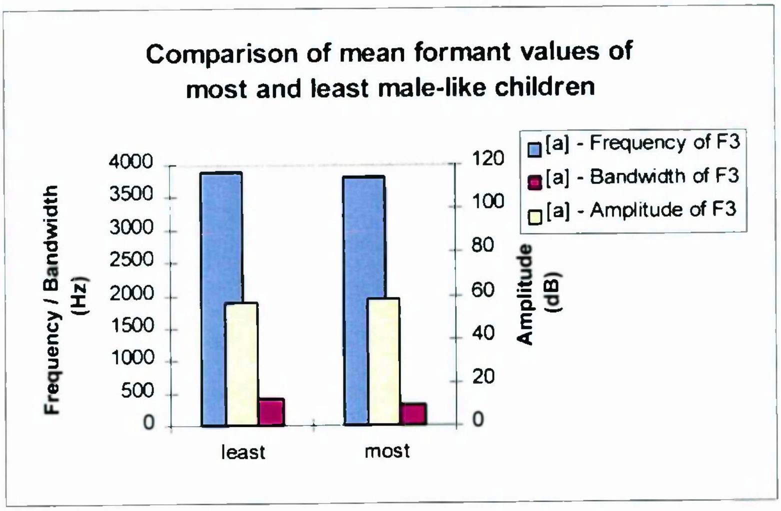


Figure A7.4 First formant information of [i] in most and least male-like children

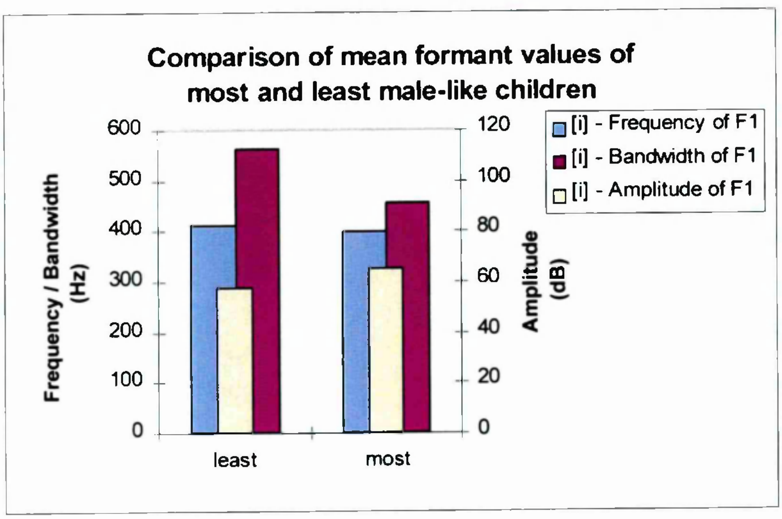


Figure A7.5 Second formant information of [i] in most and least male-like children

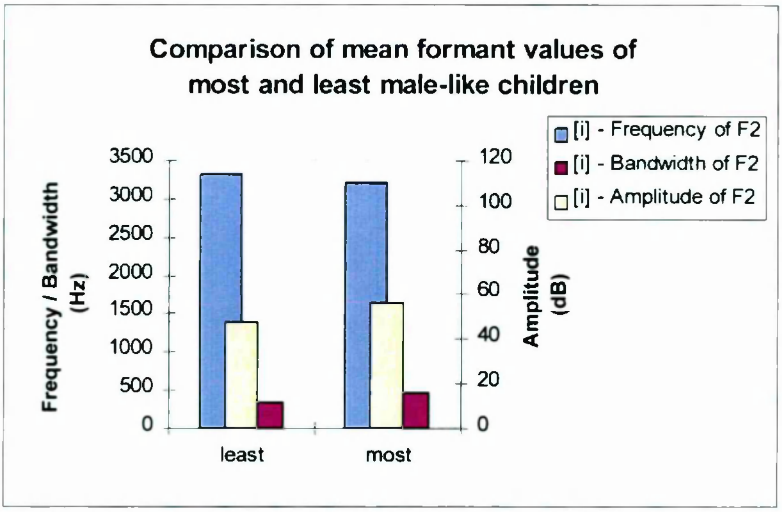


Figure A7.6 Third formant information of [i] in most and least male-like children

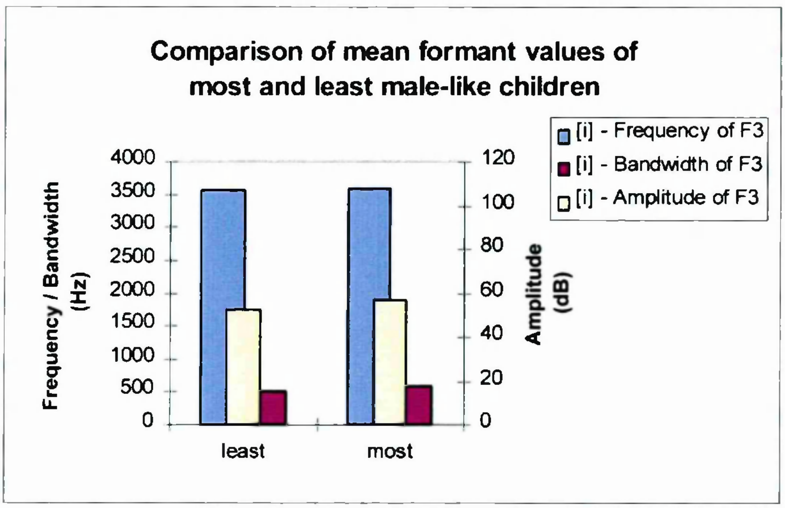


Figure A7.7 First formant information of [o] in most and least male-like children

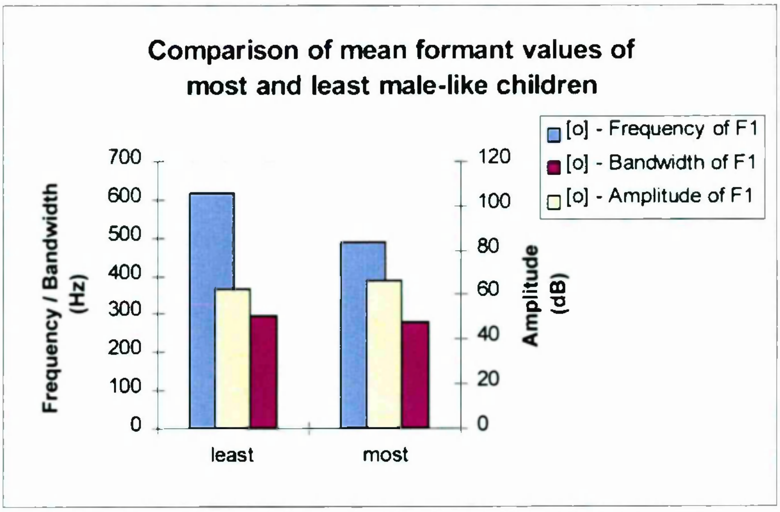


Figure A7.8 Second formant information of [o] in most and least male-like children

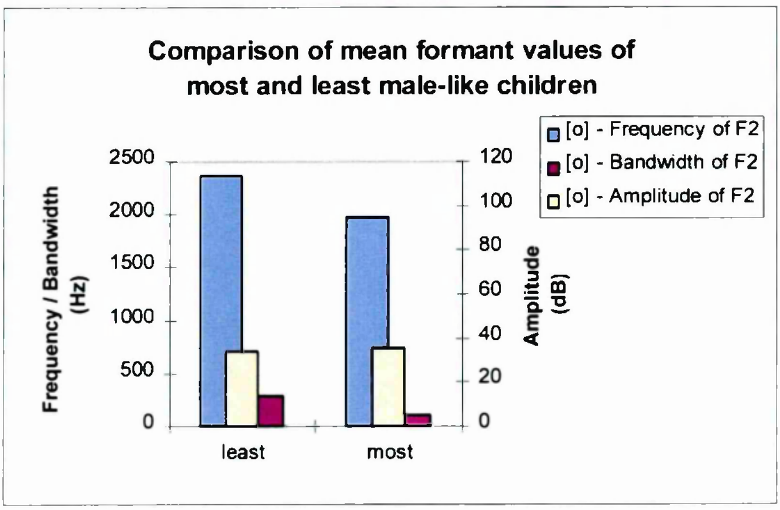
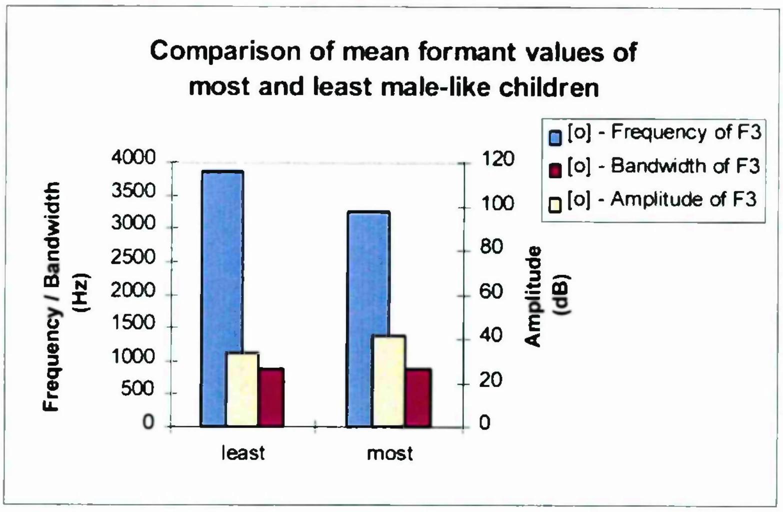


Figure A7.9 Third formant information of [o] in most and least male-like children



Appendix 8

The following children were judged to be the most or least male-like from all of the children in the study based on the vowel sample. The scores listed beside each child are the maleness scores (i.e. the number of judges who responded 'male' in the gender recognition task) for the vowel sample. Shading indicates subjects whose data was not reported in the acoustic results.

Table A8.1 Most and least male-like children

Least male-like	
Child	Score (max. 16)
fc01	0
fc05	0
fc27	0
fk13	0
fk31	0
mk29	0
fc04	1
fc07	1
fc16	1
fc17	1
fc23	1
fc37	1
fc46	1
fk22	1
mk10	1

Most male-like	
Child	Score (max. 16)
fc32	15
fk33	15
fk40	15
mc14	15
mc31	15
mc48	15
mc50	15
mc55	15
mk01	15
mk03	15
mk41	15
mk42	15

Appendix 9

Figure A9.1 Frequency distribution of mean Fo values measured from the sentence sample

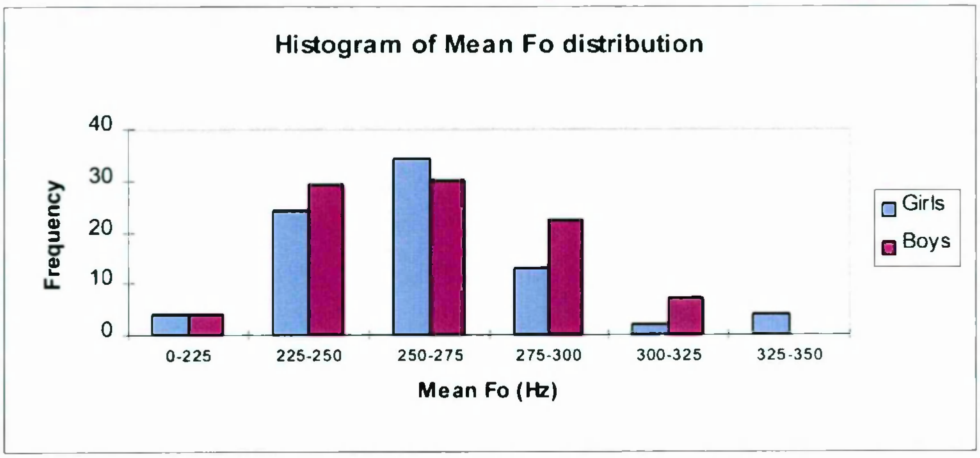


Figure A9.2 Frequency distribution of absolute jitter values measured from the sentence sample

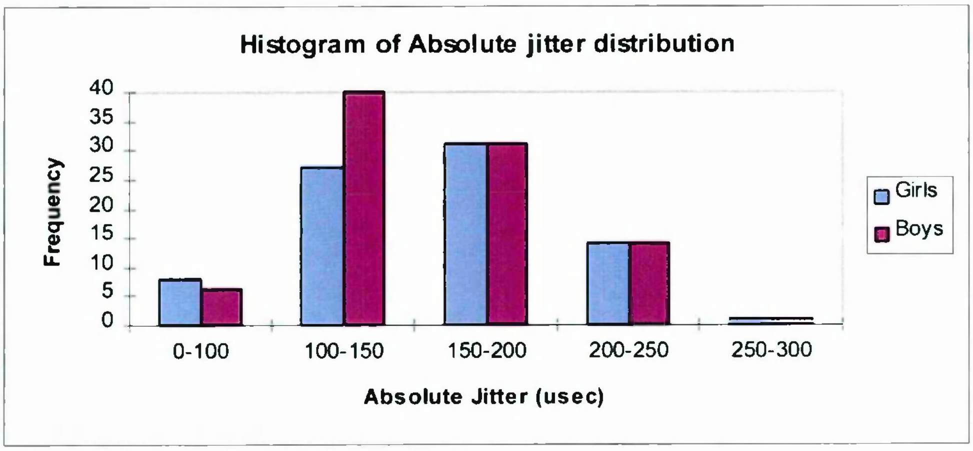
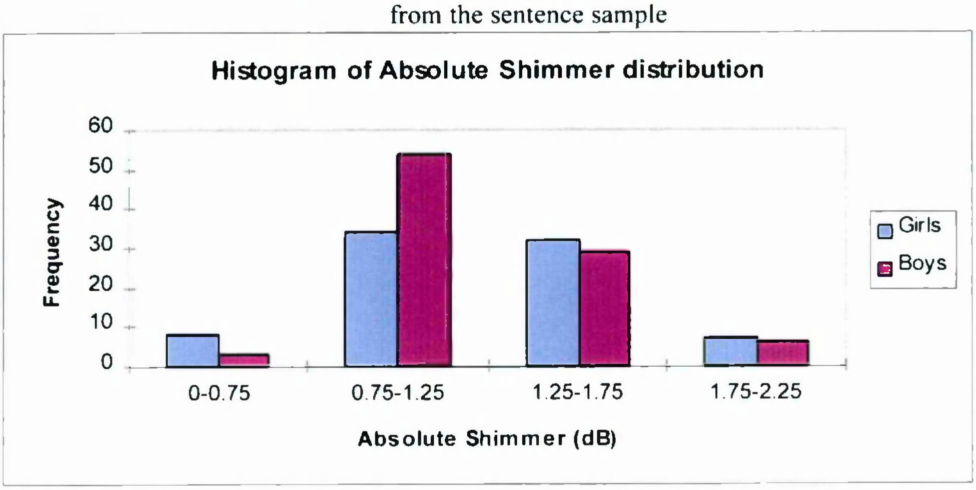


Figure A9.3 Frequency distribution of absolute shimmer values measured from the sentence sample



Appendix 10

The following references represent those papers of relevance which were published during the course of preparation of the thesis. The reader should note that the content of these publications does not diverge greatly from material presented within the thesis.

Nairn, M. (1995) The perception of gender differences in the speech of 4½ - 5½ year old children, Proceedings of the XIIIth International Congress of Phonetic Sciences, Vol. 2, pp.302-305, Stockholm

Lee, A., Hewlett, N. and Nairn, M. (1996) Voice and gender in children, In LANGUAGE AND GENDER: INTERDISCIPLINARY PERSPECTIVES, pp. 194-204, Mills, S. (editor), London: Longman Publishing

LIST OF REFERENCES

Ananthapadmanabha, T.V. (1984) Acoustic analysis of voice source dynamics. *Quarterly Progress Status Report,* **(2-3)** Speech Transmission Lab., Royal Institute of Technology, Stockholm: pp.1-24

Baken, R.J. and Daniloff, R.G. (editors) (1987) *CLINICAL MEASUREMENT OF SPEECH AND VOICE,* London: Taylor and Francis Ltd.

Beck, J.M. (1988) Organic Variation and Voice Quality, Unpublished Ph.D.Thesis, University of Edinburgh

Bennett, S. (1981) Vowel formant frequency characteristics of preadolescent males and females. *Journal of the Acoustical Society of America,* **(69),** pp.231-238

Bennett, S. (1983) A 3 year longitudinal study of school-aged children's fundamental frequencies. *Journal of Speech and Hearing Research,* **(26),** pp.137-142

Bennett, S. and Montero-Diaz, L. (1982) Children's perceptions of speakers' sex, *Journal of Phonetics,* **(10),** pp.113-121

Bennett, S. and Weinberg, B. (1979a) Acoustic correlates of perceived sexual identity in preadolescent children's voices, *Journal of the Acoustical Society of America,* **(66),** pp.989-1000

Bennett, S. and Weinberg, B. (1979b) Sexual characteristics of preadolescent children's voices. *Journal of the Acoustical Society of America,* **(65),** pp.179-189

Bernstein, I. (1978) Quotation cited in *SEX-DIFFERENCES IN HUMAN COMMUNICATION*, Eakins, B. and Eakins, R.G. (1978), Boston: Houghton Mifflin Co.

Bickley, C. (1982) Acoustic analysis and perception of breathy vowels, *M.I.T.Working Papers in Speech Communication,* **(1),** pp.71-81

Birdwhistell, R. (1970) Masculinity and Femininity as Display. In *KINESICS AND CONTEXT,* Philadelphia: University of Pennsylavania Press

Bless, D.M., Biever, D., and Shaikh, A. (1986) Comparisons of vibratory characteristics of young adult males and females, Proceedings of the *International Conference on Voice,* **(2)** pp.46-54, Kurume

Bohan, J.S. (1993) Regarding gender - essentialism, constructionism and feminist psychology, *Psychology of Women Quarterly,* **(17)**, pp.5-21

Böhme, G. and Stuchlik, G. (1995) Voice profiles and standard voice profile of untrained children, *Journal of Voice,* **(9)**, pp.304-307

Bresser, A. and Gunzburger, D. (1985) Voice recognizability of prepubertal boys and girls, *Progress Report,* **(10.2)** Institue of Phonetics, Utrecht: pp.25-32

Butler, P. (1976) *SELF ASSERTION FOR WOMEN*, San Francisco: Canfield

Childers, D.G. and Wu, K. (1991) Gender recognition from speech. Part II: fine analysis. *Journal of the Acoustical Society of America,* **(90)**, pp.1841-1856

Coates, J. (1995) Language, gender and career. In *LANGUAGE AND GENDER: INTERDISCIPLINARY PERSPECTIVES,* Mills, S. (editor), London: Longman Publishing

Coleman, R.O. (1971) Male and female voice quality and its relationship to vowel formant frequencies, *Journal of Speech and Hearing Research,* **(14)**, pp.565-577

Coleman, R.O. (1973) A comparison of the contributions of two vocal characteristics to the perception of maleness and femaleness in the voice, *Quarterly Progress Status Report,* **(2-3)** Haskins Laboratory, Stockholm

Coleman, R.O. (1976) A comparison of the contribution of two voice quality characteristics to the perception of maleness and femaleness in the voice. *Journal of Speech and Hearing Research,* **(19)**, pp.160-180

Cooley, R.N. (1964) Radiology and Radiologic Examination of the Larynx, *American Journal of Medical Science,* **(248)**, pp.601

Crawford, M. (1995) *TALKING DIFFERENCES - ON GENDER AND LANGUAGE,* 1st ed., London: Sage Publications

Crelin, E.S. (1973) *FUNCTIONAL ANATOMY OF THE NEWBORN,* New Haven: Yale University Press

Davis, S.B. (1979) Acoustic charactersitics of normal and pathological voices. In *SPEECH AND LANGUAGE: BASIC ADVANCES IN RESEARCH AND PRACTICE,* Lass, N.J. (editor), New York: Academic Press

DeWeese, D.D. and Saunders, W.H. (1973) *TEXTBOOK OF OTOLARYNGOLOGY,* St. Louis: CV Mosby

Dickson, D.R. and Dickson, W. (1971) Functional Anatomy of the Human Larynx, *Proceedings of the Pennsylvannia Academy of Opthalmology and Otolaryngology*, p.29

Dickson, D.R. and Maue-Dickson, W. (1982) *ANATOMICAL AND PHYSIOLOGICAL BASES OF SPEECH*, Boston: Little, Brown and Co.

Eakins, B. and Eakins, R.G. (1978) *SEX-DIFFERENCES IN HUMAN COMMUNICATION*, Boston: Houghton Mifflin Co.

Eguchi, S. and Hirsh, I.J. (1969) Development of speech sounds in children, *Acta Otolaryngologica*, **(Suppl. 257)**, pp.5-48

Fant, G. (1960) *ACOUSTIC THEORY OF SPEECH PRODUCTION*, 1st ed., The Hague: Mouton

Fant, G. (1966) A note on vocal tract size factors and non-uniform F-pattern scalings. *Quarterly Progress Status Report*, **(4)** Speech Transmission Lab., Royal Institute of Technology, Stockholm: pp.22-30

Fant, G. (1973) *SPEECH SOUNDS AND FEATURES*, Cambridge: M.I.T.

Fant, G. (1975) Non-uniform vowel normalisation, *Quarterly Progress Status Report*, **(2-3)** Speech Transmission Lab., Royal Institute of Technology, Stockholm: pp.1-19

Fant, G. (1979) Glottal source and excitation analysis, *Quarterly Progress Status Report*, **(4)** Speech Transmission Lab., Royal Institute of Technology, Stockholm: pp.85-107

Fichtelius, A., Johansson, I., Nordin, K. (1980) Three investigations of sex-associated speech variation in day school, *Womens Studies International Quarterly*, **(3)**, pp.219-225

Fields, S. and Dunn, F. (1973) Correlation of Echographic Visualizability of Tissue with Biological Composition and Physiological State, *Journal of the Acoustical Society of America*, **(54)**, pp.809-812

Fink, B.R., Basek, M., Epanchin, V. (1956) The Mechanism of Opening of the Human Larynx, *Laryngoscope*, **(66)**, pp.410

Gobl, C. (1988) Voice source dynamics in connected speech, *Quarterly Progress Status Report*, **(1/1988)** Speech Transmission Lab., Royal Institute of Technology, Stockholm: pp.123-159

Goerttler, K. (1950) Die Anordnung, Histologie und Histogenese Der Quergestreiften Muskulatur in Menschlichen Stimmband, *Zeitschrift für Anatomie und Entwickelungsgeschichte*, **(115)**, pp.352-401

Gunzburger, D. (1984) Perception of some male/female voice characteristics. *Progress Report,* **(9.2)** Institute of Phonetics, Utrecht: pp.15-26

Gunzburger, D., Bresser, A., Ter Keurs, M. (1987) Voice identification of prepubertal boys and girls by normally sighted and visually handicapped subjects. *Language and Speech,* **(30),** pp.47-58

Hasek, C., Singh, S., Murry, T. (1980) Acoustic attributes of preadolescent voices. *Journal of the Acoustical Society of America,* **(65),** pp.1262-1265

Henton, C.G. and Bladon, R.A.W. (1985) Breathiness in normal female speech: Inefficiency versus desirability, *Language and Communication,* **(5),** pp.221-227

Hirano, M. (1975) Phonosurgery - Basic and Clinical Investigations, *Otologia (Fukuoka),* **(21),** pp.239-440

Hirano, M. (1977) Structure and Vibratory Behaviour of the Vocal Folds. In *DYNAMIC ASPECTS OF SPEECH PRODUCTION,* Sawashima, M. and Cooper, F.S. (editors), Tokyo: University of Tokyo Press

Hirano, M. (1981) *CLINICAL EXAMINATION OF VOICE,* Vienna: Springer Verlag

Hirano, M. (1983) The structure of the vocal folds. In *VOCAL FOLD PHYSIOLOGY,* Stevens, K. and Hirano, M. (editors), Tokyo: University of Tokyo Press

Hirano, M., Koike, Y., Hirose, S., Morio, M. (1973) Structure of the Vocal Cord as a Vibrator, *Journal of Otolaryngology of Japan,* **(76),** pp.1341-1348

Hirano, M., Kurita, S., Nakashima, T. (1983) Growth, development and ageing of human vocal folds. In *VOCAL FOLD PHYSIOLOGY: CONTEMPORARY RESEARCH AND CLINICAL ISSUES,* Bless, A. and Abbs, A. (editors), San Diego: College Hill Press,Inc.

Holmberg, E., Hillman, R., Perkell, J. (1988) Glottal airflow and transglottal pressure measurements for male and female speakers in soft, normal and loud voice. *Journal of the Acoustical Society of America,* **(84),** pp.511-529

Horii, Y. (1979) Fundamental frequency perturbation observed in sustained phonation, *Journal of Speech and Hearing Research,* **(22),** pp.5-19

Horii, Y. (1980) Vocal shimmer in sustained phonation, *Journal of Speech and Hearing Research,* **(23),** pp.202-209

Hunter, W.S. and Garn, S. (1972) Disproportionate sexual dimorphism in the human face, *American Journal of Physical Anthropology,* **(36),** pp.133-138

Ingemann, F. (1968) Identification of the speaker's sex from voiceless fricatives, *Journal of the Acoustical Society of America,* **(44),** pp.1144

Ingrisano, D. and Thompson, R. (1975) Sex identification of preschool children's voices, [Unpublished]

Ingrisano, D., Weismer, G., Shuckers, G.H. (1980) Sex identification of preschool children's voices, *Folia Phoniatrica,* **(32),** pp.61-69

Kahane, J. (1978) A morphological study of the human prepubertal and pubertal larynx, *American Journal of Anatomy,* **(151),** pp.11-20

Kakita, Y., Hirano, M., Kawasaki, H., Matsushita, H. (1976) Schematical presentation of vibration of the vocal cords as a layer-structured vibrator. Normal larynges. *Journal of Otolaryngology of Japan,* **(79),** pp.1333-1340

Kakita, Y., Hirano, M., Ohmaru, K. (1981) Physical Properties of the Vocal Fold Tissue: Measurement on excised larynx. In *VOCAL FOLD PHYSIOLOGY,* Hirano, M. and Stevens, K. (editors), Tokyo: University of Tokyo Press

Kaplan, H. (1971) *ANATOMY AND PHYSIOLOGY OF SPEECH,* New York: McGraw-Hill, Inc.

Karlsson, I. and Rothenberg, M. (1992) Inter-cultural variations in gender-based language differences in young children, *Quarterly Progress Status Report,* **(1)** Speech Transmission Lab., Royal Institute of Technology, Stockholm: pp.1-17

Kent, R.D. (1976) Anatomical and neuromuscular maturation of the speech mechanism: evidence from acoustic studies. *Journal of Speech and Hearing Research,* **(19),** pp.421-447

Kent, R.D. and Forner, L. (1979) Developmental study of vowel formant frequencies in an imitation task. *Journal of the Acoustical Society of America,* **(65),** pp.208-217

Kent, R.D. and Read, C. (1992) *THE ACOUSTIC ANALYSIS OF SPEECH,* San Diego: Singular Publishing Inc.

King, E. (1952) A roentgenographic study of pharyngeal growth, *Angle Orthod.* **(22),** pp.23-37

Kirchner, J.A. (1970) *PHYSIOLOGY OF THE LARYNX,* Rochester, Minn.

Klatt, D.H. and Klatt, L.C. (1990) Analysis, synthesis and perception of voice quality variations among male and female talkers, *Journal of the Acoustical Society of America,* **(87),** pp.820-857

Kurita, S. (1980) Layer Structure of the Human Vocal Fold - Morphological investigations, *Otologia (Fukuoka),* **(26),** pp.993-997

Ladefoged, P. (1982) *A COURSE IN PHONETICS,* 2nd ed., New York: Harcourt Brace Jovanovich

Ladefoged, P. and Antoñanzas-Barroso, N. (1985) Computer measures of breathy voice quality, *U.C.L.A. Working Papers in Phonetics,* **(61),** pp.79-86

Ladefoged, P. and Broadbent, D. (1957) Information conveyed by vowels, *Journal of the Acoustical Society of America,* **(29),** pp.98-104

Ladefoged, P., Maddieson, I., Jackson, M., Huffman, M.K. (1987) Characteristics of the voice source, *U.C.L.A. Working Papers in Phonetics,* **(67),** pp.119-125

Lakoff, R. (1978) Women's language. In *WOMEN'S LANGUAGE AND STYLE,* Butturff, D. and Epstein, E.L. (editors), Akron, Ohio: University of Akron

Lass, N.J., Hughes, K.R., Bowyer, M.D., Waters, L.T., Bourne, V.T. (1976) Speaker sex identification from voiced, whispered and filtered isolated vowels, *Journal of the Acoustical Society of America,* **(59),** pp.675-678

Lehiste, I. and Meltzer, D. (1973) Vowel and Speaker Identification in natural and synthetic speech. *Language and Speech,* **(16),** pp.356-364

Lehiste, I. and Peterson, G.E. (1961) Some basic considerations in the analysis of intonation, *Journal of the Acoustical Society of America,* **(33),** pp.419-425

Lindblom, B. (1972) Comments on Paper 15, "Development of speech sounds in children", Eguchi, S. and Hirsh, I.J. Proceedings of the *International Symposium on Speech Communication Ability and Profound Deafness,* pp.159-162, Washington D.C. Alexander Graham Bell Association for the Deaf

Lindblom, B. and Sundberg, J. (1971) Acoustical consequences of lip, tongue, jaw and larynx movements, *Journal of the Acoustical Society of America,* **(50),** pp.1166-1179

Luchsinger, R. and Arnold, G.E. (1965) *VOICE - SPEECH - LANGUAGE,* Belmont, CA: Wadsworth

Maltz, D. and Borker, R. (1982) A cultural approach to male female miscommunication. In *LANGUAGE AND SOCIAL IDENTITY,* Gumperz, J. (editor), Cambridge: Cambridge University Press

Marshall, G. (1972) Sex typing of speech of prepubertal children, Unpublished Ph.D.Thesis, Louisiana State University

Mattingly, I.M. (1966) Speaker variation and vocal tract size, [Abstract] *J. Acou. Soc. Am.* **(39)**, pp.1219

Maue, W. (1970) Cartilages, Ligaments and Articulations of the Adult Human Larynx, Unpublished Ph.D.Thesis, University of Pittsburgh

Maue, W.M. and Dickson, D.R. (1971) Cartilages and Ligaments of the Adult Human Larynx, *Archives of Otolaryngology,* **(94)**, pp.432

McCarthy, D. (1953) Some Possible Explanations of Sex Differences in Language Developments and Disorders, *Journal of Psychology*, **(35)**, pp. 155-160

McNicol, D. (1972) *A PRIMER OF SIGNAL DETECTION THEORY,* London: Allen & Unwin

Mullenix, J.W., Johnson, K.A., Topcu-Gurdun, M. and Farsworth, L.M. (1995) The perceptual representation of voice gender, *Journal of the Acoustical Society of America,* **(98)**, pp.3080-3095

Murphy, A.T. (1964) *FUNCTIONAL VOICE DISORDERS,* Englewood Cliffs, N.J. Prentice-Hall

Negus, V.E. (1962) *THE COMPARATIVE ANATOMY AND PHYSIOLOGY OF THE LARYNX,* 5th ed., New York: Hafner

Ni Chasaide, A. and Gobl, C. (1997) Voice Source Variation. In *HANDBOOK OF PHONETIC SCIENCES,* Laver, J. and Hardcastle, W. (editors), London: Blackwell, pp. 427-461

Ohala, J.J. (1983) Cross-language use of pitch: an ethological view, *Phonetica,* **(40)**, pp.1-18

Okada, M. (1978) An investigation of the blood vessels of the vocal fold. *Otologia (Fukuoka),* **(24)**, pp.974-994

Orlikoff, R. (1990) Assessment of the dynamics of vocal fold contact from the electroglottogram: Data from normal male subjects, [Unpublished] cited in Gender comparisons of children's vocal fold contact behaviour, Robb, M.P. and Simmons, J.O., *Journal of the Acoustical Society of America,* **(88)**, pp.1318-1322, (1990)

Peterson, G.E. and Barney, H.L. (1952) Control methods used in the study of vowels. *Journal of the Acoustical Society of America,* **(24)**, pp.175-184

Pick, T.P. and Howden, R. (eds.), (1977) *GRAY'S ANATOMY,* 15th ed., New York: Bounty Books

Rabiner, L.R. and Schafer, R.W. (1978) *DIGITAL PROCESSING OF SPEECH SIGNALS,* Englewood Cliffs, New Jersey: Prentice-Hall

Renfrew, C.E. (1969) The Bus Story - A test of continuous speech, North Place, Old Headington, Oxford: C.E. Renfrew

Robb, M.P. and Simmons, J.O. (1990) Gender comparisons of children's vocal fold contact behaviour. *Journal of the Acoustical Society of America,* **(88),** pp.1318-1322

Romanes, G.J. (1971) *HEAD, NECK AND BRAIN,* London: Oxford University Press

Sachs, J. (1975) Cues to the identification of sex in children's speech. In *LANGUAGE AND SEX: DIFFERENCE AND DOMINANCE,* Thorne, B. and Henley, N. (editors), Rowley, Mass. Newbury House

Sachs, J., Lieberman, P., Erickson, D. (1973) Anatomical and cultural determinants of male and female speech. In *LANGUAGE ATTITUDES: CURRENT TRENDS AND PROSPECTS,* Shuy, R.W. and Fasold, R.W. (editors), Washington D.C. Georgetown Univ. Press

Samuels, M.L. (1989) *STATISTICS FOR THE LIFE SCIENCES,* San Francisco: Dellen Publishing Co.

Schwartz, M.F. (1968) Identification of speaker sex from isolated voiceless fricatives. *Journal of the Acoustical Society of America,* **(43),** pp.1178-1179

Schwartz, M.F. and Rine, H. (1968) Identification of speaker sex from isolated whispered vowels, *Journal of the Acoustical Society of America,* **(44),** pp.1736-1737

Sinclair, D. (1978) *HUMAN GROWTH AFTER BIRTH,* 3rd ed., London: Oxford University Press

Sorenson, D.N. (1989) A fundamental frequency investigation of children aged 6-10 years old. *Journal of Communication Disorders,* **(22),** pp.115-123

Stevens, K. and House, A. (1955) Development of a quantitative description of vowel articulation, *Journal of the Acoustical Society of America,* **(27),** pp.484-493

Takahashi, H. and Koike, Y. (1975) Some perceptual dimensions and acoustical correlates of pathological voices, *Acta Otolaryngol. (Suppl.),* **(338),** pp.1-24

Tanner, J.M. (1978) *FOETUS INTO MAN - PHYSICAL GROWTH FROM CONCEPTION TO MATURITY,* London: Open Books Publishing Ltd.

Temple, R.A.M. (1988) Sex-specific aspects of the voicing contrast in French stop consonants. Unpublished M.PhilThesis, Laboratory of Phonetics, University of Oxford

Templin, M. (1957) *CERTAIN LANGUAGE SKILLS IN CHILDREN: THEIR DEVELOPMENT AND INTERRELATIONSHIP*, Minneapolis: University of Minnesota Press

Titze, I.R. (1989) Physiologic and acoustic differences between male and female voices, *Journal of the Acoustical Society of America,* **(85)**, pp.1699-1707

Tonkova-Yampol 'skaya, R.V. (1973) Razvitie rechevoi intonatsii u detei pervykh dvukh let zhini (the development of speech intonation in children during the first two years of life). In *STUDIES OF CHILD LANGUAGE DEVELOPMENT,* Ferguson, C. and Slobin, D. (editors), New York: Holt, Rinehart, Winston

Trudgill, P. (1983) *SOCIOLINGUISTICS: AN INTRODUCTION TO LANGUAGE AND SOCIETY,* Revised ed., London: Penguin

Vuorenkoski, V. (1978) Fundamental voice frequency during abnormal growth, and after androgen treatment, *Archives of Disease of Childhood,* **(53)**, pp.201-209

Walker, G.F. and Kowalski, C.J. (1972) On the growth of the mandible, *American Journal of Physical Anthropology,* **(36)**, pp.111-117

Watson, I.M.C. (1987) Problems in quantifying the child voicing contrast, [Unpublished] cited in Sex-specific aspects of the voicing contrast in French stop consonants, Temple, R.A.M. [Unpublished M.PhilThesis, Laboratory of Phonetics, University of Oxford]

Weinberg, B. and Bennett, S. (1971) Speaker sex recognition of 5 and 6 year old children's speech. *Journal of the Acoustical Society of America,* **(50)**, pp.1210-1213

Woods, N. (1992) It's not what she says, it's the way that she says it: the influence of speaker-sex on pitch and intonational patterns. *Progress Report from Oxford Phonetics,* **(5)**, pp. 84-95, Oxford, OUPL

Wrench, A.A. (1995) Analysis of fricatives using multiple centres of gravity, *Proceedings of the XIIIth International Congress of Phonetic Sciences (ICPhS),* **(4)** pp.460-463, Stockholm: ICPhS 95

Wrench, A.A., Watson, J.M.M., Soutar, D.S., Robertson, A.G., and Laver, J. (1994) Fast formant estimation of children's speech, *Proceedings of the International Conference on Spoken Language Processing (ICSLP),* **(3)** pp.1651-1654, Yokohama: The Acoustical Society of Japan

Wu, K. and Childers, D.G. (1991) Gender recognition from speech:Part 1: Coarse analysis, *Journal of the Acoustical Society of America,* **(90)**, pp.1828-1840